PB85-205607

Empirical Test of Economic Model for Innovation in the Chemical Industry

Meta Systems, Inc., Cambridge, MA

Prepared for

Environmental Protection Agency, Washington, DC

Dec 84

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-205607

Empirical Test of Economic Model for Innovation in the Chemical Industry

Prepared by:

Meta Systems Inc
10 Holworthy Street
Cambridge, MA 02138

Contract: 68-01-6674

Edward B. Leviton, Project Officer
Economics and Technology Division
Office of Pesticides and Toxic Substances
U.S. Environmental Protection Agency
Washington, D.C. 20460

December 1984

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

50272-101

REPORT DOCUMENTATION PAGE

1. REPORT NO. EPA/4-85-002

2.

3. Recipient's Accession No. PB85 205607/AS

4. Title and Subtitle

Empirical Test of Economic MOdel for
Innovation in the Chemical Industry

5. Report Date December 1984

6.

7. Author(s)

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

Meta Systems Inc.
10 Holworthy Street
Cambridge, MA 02138

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.

(C) 68-01-6674
(G)

12. Sponsoring Organization Name and Address

Economics and Technology Division
Office of Toxic Substances
U.S. Environmental Protection Agency
Washington, DC 20460

13. Type of Report & Period Covered

14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

The object of this report was to develop a model describing the innovation process in the chemical industry. The model developed provides a means of measuring and analyzing the impacts on innovation of various regulations. Of particular interest is the Toxic Substance Control Act(TSCA), which many critics argue has had a detrimental impact on innovation in the Chemical industry.

First, a model of the innovation process is developed. This model contains three equations, which are also estimated using available data. The model works well explaining a great deal about the innovation process. Second, an attempt was made to analyze the imapcts of TSCA. The data needed to complete a complete analysis of TSCA.

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement

19. Security Class (This Report)

20. Security Class (This Page)

21. No. of Pages 191

22. Price

(See ANSI-Z39.18)

See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

DISCLAIMER

This report was prepared under contract to an agency of the U.S. Government. Neither the U.S. Government nor any of its employees, contractors, subcontractors or their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party use of any information disclosed in this report.

Publication of the data in this document does not signify that the contents necessarily reflect the joint or separate views and policies of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGEMENTS

This study was conducted by Meta Systems Inc, under the direction of Dr. Peneolope H. Schafer and Dr. Peter J. Morgan. However, as is often the case with projects of this type, this report is the result of work by several people. I would like to acknowledge these efforts and thank the participants for their contributions. Much of the empirical work dealing with the single innovation measure (all patents) and with the U.S. Federal Trade Commission Line of Business data was done by Dr. Morgan with the able assistance of Sharmin Mossavar-Rahmani. We thank Dr. David F. Lean, Deputy Manager, and the rest of the Line of Business Program staff, for making their data available and accessible to us. The conceptual and empirical work dealing with the dual measures of innovative (distinguishing between product and process patents) was done by Professor Mark A. Schankerman of New York University and Adam B. Jaffe. They also provided invaluable assistance and insights into other areas of the report, particularly Professor Schankerman's contributions to the overall conceptual framework used in this analysis.

We received very thorough and valuable comments from three outside readers: Professors Zvi Griliches of Harvard University, C.A. Knox Lovell of University of North Carolina at Chapel Hill, and Ariel Pakes, of the University of Wisconsin at Madison. They, however, bear no responsibility for the contents of this report.

We would also like to thank Edward B. Leviton, Sammy Ng, and Michael Shapiro, of the Economics and Technology Division, Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency, for their assistance and guidance throughout this analysis.

The report would not be available to readers if it were not for the efforts of Kathleen R. West, Jeanne F. Quan, Andrew Bell, and Charlotte Young.

Penelope H. Schafer
Project Director

Table of Contents

Page Number

Chapter 1: Executive Summary

Introduction 1-1
Toxic Substances Control Act (TSCA) 1-1
Research Approach and Data Sources 1-3
Knowledge Production Function 1-6
Revenue Equation 1-7
R&D Demand Function 1-8
Using the Model to Analyze the Impacts of TSCA 1-9
Appendix 1-A: Data Required by Premanufacturing Notice (PMN) . . 1A-1

Chapter 2: Research Approach and Methodology

Introduction 2-1
Approaches Considered 2-1
Conceptual Approach Adopted 2-3
Definitions and Sources of Data 2-10
Principal Econometric Problems 2-17
Appendix 2A: Characteristics of Firms. 2A-1
Appendix 2B: Test of Reliability of Patent Titles as a Means to Classify Patents 2B-1

Chapter 3: Knowledge Production Function

Introduction 3-1
General Model 3-1
Estimates Based on Aggregate Patent Data 3-5
Estimates Based on the FTC Line of Business Data 3-15
Estimates that Distinguish Between Product and Process Innovation 3-21
Appendix 3A: Results Including Zero Patent and R&D Observations, and Measurements of Scale Effects 3A-1

Chapter 4: Revenue Function

Introduction 4-1
Estimates Based on Firm-Level Data 4-4
Estimates Based on FTC Line of Business Data 4-15
Estimates that Distinguish Between Product and Process Innovations 4-23
Appendix 4A: Estimates Including Scale Effects, and Zero Patent and R&D Observations. 4A-1
Appendix 4B: Firm-Specific Propensity to Patent and the Revenue Function 4B-1

Table of Contents
(continued)

Page Number

Chapter 5: R&D Demand Function
Introduction 5-1
Estimates Based on Firm-Level Data 5-2
A Model that Distinguishes Between Product and Process Innovations 5-7
Appendix 5A: Derivation of the R&D Equation 5A-1

Chapter 6: Interpretation of Results
Introduction 6-1
Interpretation of Results 6-2
Main Lines for Further Research 6-6
Use of Model to Assess Impacts of TSCA 6-11

References

Chapter 1

Executive Summary

Introduction

The objective of the research presented in this report was to develop a model describing the innovation process in the chemical industry. The model developed here provides a means of measuring and analyzing the impacts on innovation of various regulations. Of particular interest is the Toxic Substances Control Act (TSCA), which many critics argue has had a detrimental impact on innovation in the chemical industry.

The research effort falls into four major tasks. The results of these tasks are described briefly in this chapter, and are presented in detail in the following chapters. First, a model of the innovation process was constructed. This model contains three equations, which describe the major steps in the process. Second, data were collected so that the three relationships in the model could be estimated. The details of these two tasks are described in Chapter 2. Third, the model was tested, using the data collected. Chapters 3, 4, and 5 present the details of the estimation work for each of the three equations. The model works very well, explaining a great deal about the innovation process. The fourth task was to describe how the model can be used to analyze the impacts of TSCA. This is done in Chapter 6. Since the data necessary to perform the impact analysis were not available to the authors at this time, the impact analysis itself was not undertaken.

Toxic Substances Control Act (TSCA)

TSCA became effective on January 1, 1977 and gave the U.S. Environmental Protection Agency (EPA) broad authority to monitor and regulate chemicals from product development and production through disposal. Since July 1, 1979, manufacturers and importers of new chemicals are required to file a Premanufacture Notice (PMN) with EPA at least 90 days before commencement of manufacture or importation*. This enables EPA to review the new chemical to determine if any controls or limitations on its manufacture, use, or disposal are necessary. An exact count of the number and disposition of PMN's filed to date is not available at this time. After a slow start, however, the annual rate of filing PMN's has risen to a current level of about 100 per month.

Based on the PMN and other available information, EPA can take the following actions:

- Accept chemicals without further consideration, i.e. no regulations or restrictions.
- Select substances for more detailed evaluations and for which additional exposure or toxicological data are needed.
- Identify chemicals that will be subject to requirements for followup reporting during their commercial development.
- Select a limited number for immediate regulatory attention, such as limiting production and/or use, including a ban on production.

As a result TSCA, and more particularly the PMN requirements, impose a series of costs on manufacturers and importers of new chemicals.** Some of these are direct costs, such as the time and expense of compiling and submitting the data required by the PMN and the potentially higher manufacturing costs resulting from rulings by EPA. More important are the numerous indirect costs, such as the potential delay in introducing new chemicals, fear of trade secret disclosures, possible restrictions on the ability to market the chemical, and the uncertainty over the direct and indirect costs.

---

* A list of the data required by the PMN can be found in Appendix 1-A.

** Under TSCA, EPA also reviews existing chemicals to determine what, if any, restrictions should be placed on their manufacturing, use and/or disposal. Currently, EPA is reviewing 5 to 10 existing chemicals annually.

The chemical industry has argued that TSCA will result in a significant decrease in the amount of innovation occurring. They argue that the costs and the uncertainties resulting from TSCA will mean that new products are less profitable and that profits are less sure; thus fewer new chemicals will be developed. However, many factors in addition to TSCA, both macroeconomic and firm specific, affect the decision of how much a firm will spend on R&D and in what areas to concentrate their R&D efforts. Therefore, the impacts of TSCA can be analyzed only if these other influences are taken into account. The model developed in this report does take these other influences into account, allowing for an analysis of the impacts of TSCA itself.

Research Approach and Data Sources*

The model focuses on the firm's decision process and assumes that firms are profit maximizers. This process has three main steps and can be thought of as a loop, with the results of each step feeding into the follow steps. First, the firm chooses a level of R&D effort, based on their sales level, expected growth rate, and knowledge stock. Second, this R&D effort results in the production of new knowledge (measured in terms of patents). Third, this new knowledge increases the revenues derived from a given level of capital and labor inputs. This is because the patent is either for a new, improved product that sells for a higher price, or for a new process that produces the same product with less labor and/or capital.

---

* See Chapter 2 for a more detailed discussion.

There are two versions of the model. One employs a single measure of innovation--all patents. However, TSCA is expected to differentially impact product and process innovation since firms are required to file PMN's for new products, but not for new processes. Therefore, a second version of the model was developed that has two measures of innovation--product patents and process patents. This distinction allows for the differentiated impacts.

The decision process proceeds temporarily in the order described above. However the first relationship, describing the decision about the level of R&D (the R&D Demand Equation), is dependent on the second and third relationships. The second and third equations describe technological relationships. In other words, they describe the amount of output resulting from each level and combination of relevant inputs. As such, they do not embody any assumptions about the motivations of the firm. The second relationship, the Knowledge Production Function, describes the amount of new knowledge that results from a given level of R&D expenditures. The third relationship, the Revenue Equation, gives the amount of revenues the firm will receive from a given amount of capital, labor and knowledge inputs. In both cases, the stock of knowledge is measures in terms of number of patents.

The first equation, the R&D Demand Function, gives the amount the firm will spend on R&D given the firm's sales, expected growth rate, and its stock of knowledge. This equation is derived from the other two and embodies the behavioral assumption that the firm is a profit maximizer. As such, the firm will invest resources in R&D until the long-run marginal benefits derived from the investment equals the marginal cost of the investment.

To summarize, the three equations that comprise the model of innovation are:

1. Knowledge Production Function -- relates R&D expenditures to measure of new knowledge (patents) which can be thought of intermediate outputs.
2. Revenue Function* -- links production inputs (capital, labor and knowledge stock) to revenues.
3. R&D Demand Equation -- level of R&D expenditures depend on firms' expectations of future revenue growth, which results from increases in the stock of knowledge via the other two equations.

There are two major types of data used in the analysis: 1) financial data, and 2) measures of innovation, in this case, counts of patents. Firm-level financial data including sales revenue, R&D expenditures, capital expenditures and employment, were collected from 10-K and annual reports. Data availability and involvement in the chemical industry determined the selection of firms in the sample. In addition, similar financial data on a line of business basis were obtained from the U.S. Federal Trade Commission, Line of Busness Program. Value-added data, by four digit SIC group, were obtained from the U.S. Census of Manufactures. Counts of chemical patent applications are from the U.S. Patent Office, Office of Technology Assessment and Forecasting (OTAF). All three of the equations were estimated using the firm level financial and patent data. In addition, the Knowledge Production Function (KPF) and the Revenue Equation were estimated using the FTC Line of Business (LB) financial data and patent data matched to the LB data. The time series available was too short to allow for estimating the R&D Demand Function for the LB data. Since the PMN requirement applies to new products, but not to new processes, it may differentially affect the cost of product and process innovation. To allow for this, the three equations in the model were re-specified to include product and process patents separately. The re-specified KPF and Revenue Function were estimated, while the R&D Equation is discussed.

---

* This is not the type of revenue function in which revenues are a function of input prices and output prices.

Knowledge Production Function*

The R&D process may be conceived of as the use of various scientific and technical inputs to develop a body of technical knowledge which the firm uses to maintain and enhance its production possibilities. The Knowledge Production Function (KPF) summarizes the productivity of R&D in terms of generating new knowledge. Patents are used as the indicator of new knowledge. In other words, the KPF specifies that new knowledge (patents) in any given year are a function of the R&D stock in that year. In turn, the R&D stock is a function of R&D expenditures in past years, discounted by an appropriate R&D depreciation or decay rate.

The analysis finds a strong, stable relationship between patents and R&D. It is not possible to identify the exact length of the time lag between R&D and patents, since the various alternatives tried all yielded nearly identical results. In addition, the relationship is stronger when estimated across firms than it is over time for particular firms. In other workds, over time R&D expenditure levels and number of patents for any firm change very little and so it is difficult to distinguish the relationship. However, levels do differ among firms and thus the relationship can be estimated by comparing firms (known as estimating across firms).

---

* See Chapter 3 for a more detailed discussion.

The main findings can be compared in terms of the implied marginal products and rates of return. The empirical estimates yield very plausible implications. Using the average value for each of the variables, the estimated KPF says that an increase in the level of R&D stock of one million dollars will produce approximately one-half of an additional patent during the first year and will eventually produce a total of 2.5 patents. Therefore, the marginal cost of an additional patent is $400,000.

In order to capture the differential impacts on the cost of product and process innovation, a version of the Knowledge Production Function was estimated that had two outputs: process knowledge (process patents) and product knowledge (product patents), but aggregate R&D for the firm. TSCA is expected to increase the cost of product innovation but not the cost of process innovation. Having two outputs allows for these different impacts. On average, product patents are found to cost about one-half as much as process patents.

Revenue Equation*

The second equation in the model of innovation captures the effects of innovation on productivity, either through cost reduction or product improvement, and thus innovation's effects on the firm's revenues. Treating the knowledge stock as one of the inputs is equivalent to assuming that a firm can produce the same revenues with fewer other inputs, if it has a greater stock of knowledge. Like the KPF, the Revenue Equation does not depend on any behavioral assumptions about the actions of firms. It simply estimates the amount of revenues a firm will receive for a given combination of inputs, including the firm's stock of knowledge.

---

* See Chapter 4 for a more detailed discussion.

The analysis finds that knowledge has a significant effect on revenues, whether it is measured in terms of an R&D stock or a patent stock. There is a lag between the initial R&D expenditures for a project and the eventual impact on sales. However, this analysis was unable to specify the exact length and structure of that lag.

Again, the main findings can be interpreted in terms of the implied marginal products. Using the average value for each variable, the Revenue Equation says that in the short-run, an increase in the capital stock of one million dollars increases revenues by 0.43 million dollars, while an increase in the patent stock of one million dollars increases revenues by 1.12 million dollars. Likewise, an increase in the patent stock of one patent generates about 0.34 million dollars.

As with the KPF, the revenue equation is expanded to include a knowledge stock with two components: process and product. In this way the relative value of product and process patents can be measured. Product patents are found to be worth about three-quarters as much as process patents.

R&D Demand Function*

The third equation in the model of innovation is an R&D Demand Function. This equation embodies the behavioral assumption that firms' are long-run profit maximizers when choosing their level of R&D expenditures. In other words, they increase their R&D until the long-run marginal benefit derived from R&D equals the marginal cost. A firm's optimal level of R&D expenditures is a function of its sales and expected growth rate. It is assumed that firms adjust their R&D slowly towards this optimal level, in what is called a partial adjustment model.

---

* See Chapter 5 for a more detailed discussion.

This equation closes the feedback loop to the other two equations. It specifies that R&D expenditures depend on the firm's expectations of future sales growth resulting from increases in the stock of knowledge via the other two equations. The estimates can be used to predict firms' behavioral response to the regulation.

Using the Model to Analyze the Impacts of TSCA

There are several ways in which this model can be used to assess the impacts of TSCA. The most operational way uses the relationships as estimated in Chapters 3 through 5, along with information on the implementation of TSCA and the degree to which TSCA affects certain variables in the model.

The model relates measures of activities that are of interest to policy makers (endogenous variables such as a firm's revenues, R&D activity, and patents) to measures of other activities (exogenous variables such as the gestation lag, expected growth of sales, and the relative value or cost of product patents vs. process patents). Some of these exogenous variables, or levers, will be influenced by the implementation of the TSCA regulations. Once the affects of TSCA on these levers are identified, the impact of such changes on the endogenous variables of interest can be evaluated by simulations.

The principal levers of interest in the models are:

1. parameters of the KPF that determine the marginal cost of product and/or process patents,

2. expected growth rate of product demand ("g_s" in the R&D demand equation),

3. gestation lag ("T" in the R&D demand equation),

4. depreciation rate of patents or knowledge ("δ" in the R&D demand equation), and

5. shadow price of product patents.

The types of information, or informed guesses, needed include: number of PMN's filed; actual time between filing and EPA's ruling; disposition (including withdrawal by company); amount of information, additional to what company would generate otherwise, required by TSCA and its associated costs; and when in the development process firms file their PMN's. Using information like this, the model can be used to simulate the impacts of TSCA.

1-A

Appendix 1-A.

Data Required by
Premanufacturing Notice (PMN)

The PMN requires information on the chemical itself and on possible human exposure to, and environmental release of, the chemical. In addition, it requires the substantiation of any claims of confidentiality. Specifically, the form asks for:

- Manufacturer's name,
- Chemical identity, including molecular formula and each associated impurity,
- Production volume,
- Uses of chemical,
- Methods of transport,
- Risk assessment,
- Methods of detection in environment,
- Expected occupational exposure during manufacturing, processing, use and disposal,
- Amount of substance to be released each year into the air, water, and land due to disposal.

Chapter 2

Research Approach and Data Sources

Introduction

The analysis presented in this report is an extension and refinement of earlier work done for EPA by Meta Systems in 1982.* As such, it maintains the conceptual approach developed in the earlier study. This chapter reviews the approaches considered in the earlier work and discusses the conceptual approach adopted. It also summarizes the data used in the analysis and describes the major econometric problems encountered. Further discussions of the various functional forms estimated, and the specific samples used, are presented in the following three chapters.

Approaches Considered

Prior to selecting the innovation model described below, the literature analyzing the impacts of TSCA and other similar regulations, specifically the 1962 Kefauver-Harris amendments to pharmaceutical industry regulations, was surveyed and the alternative methodological approaches evaluated. The literature contains two major approaches to modeling innovation: 1) case studies, and 2) aggregate analyses of variables such as R&D expenditure and the production of new chemical entities at the company or 4-digit SIC level.

The case study approach is best represented by the rate of return methodology pioneered by Edwin Mansfield and others to measure the private and social returns of individual R&D projects. In theory, the approach is to calculate the return on the project, based on all identifiable costs and benefits associated with the project. In practice, this procedure require massive amounts of confidential data. Thus the sample available to researchers is very dependent on the willingness of firms to cooperate. Using this approach to estimate impacts for the industry as a whole entails further problems. Even if data for a large sample of case studies could be obtained, it would be difficult to determine whether or not the sample was representative of the industry and how to extrapolate an analysis of TSCA impacts for this sample to the industry as a whole.

---

* Meta (1982). All citations in the text are given in abbreviated form. Refer to the Reference list at the end of the report for complete citations.

Aggregative approaches to the analysis of innovation are more varied. The 1962 Kefauver-Harris amendments to regulations covering the pharmaceutical industry may be the regulatory "event" which provides the closest precedent for TSCA regulations. Although there are important differences between the two regulations, it was hoped that the literature assessing the impact of the 1962 amendments would be a fruitful source of approaches for this study. The drug studies do highlight some areas of interest to TSCA, but the majority of the literature is not very useful. Generally, the drug studies are highly data constrained, post hoc, single equation analyses dealing with very different regulatory requirements. Nevertheless, they provide a helpful starting point by focusing on the relationships between R&D expenditures, measures of innovation such as new chemical entities, demand-side factors such as expected sales growth, and supply-side factors such as the depletion of research opportunities. The drug studies highlight the difficulty of associating a particular impact with a particular government regulation, even though the event itself is quite dramatic.

Another aggregative method was proposed by a subcontractor on the 1982 study.* They proposed that a single-equation "reduced form" model be estimated, which relates a measure of innovation, such as patents or new chemical entities, to variables which affect either the demand or supply of innovation, including research factor prices, the degree of market concentration, the profitability and growth prospects of the chemicals industry, and the research intensity of the industry. Although this single-equation approach shares many features with the methodology eventually used, it was not adopted because of its lack of differentiation among different phases of the innovation process and consequent difficulties of interpretation.

Conceptual Approach Adopted

In the methodology adopted, the innovation process is viewed as having three main steps:

1. the firm's decision to invest resources in innovation (the R&D Demand Equation),
2. the process by which these resources produce new knowledge (the Knowledge Production Function or KPF),
3. the use of new knowledge to increase productivity, either through its incorporation in the production process to reduce costs, or in the product to improve its quality (the Revenue Function).

Although the innovation process unfolds temporally in this order, in terms of causation the R&D Demand Equation is derived from the other two. Embodied in the Knowledge Production Function is the cost of knowledge production. The Revenue Equation captures the gross return to new knowledge. Therefore, the parameters of these equations enter into the firm's R&D decision.

---

* Charles River Associates. See the discussion in Meta (1982).

This approach was chosen because it presents a plausible structuring of the relationship between R&D activities, the outputs of these activities in terms of their effect on factor productivity (thereby affecting sales, profits and other indicators of business activity), the feedback of these effects on the demand for R&D, and other factors which determine the costs of R&D. It is based on earlier work of Griliches, Schankerman, Pakes, and others.* Figure 2-1 is a schematic model of the firm's research and production process which underlies the three equations.

Knowledge Production Function

The left side of Figure 2-1 shows the relation between R&D inputs and the creation of an "increment to knowledge" such as a new chemical entity or new process. Increments to knowledge are not directly observable; however, patents, publications, or other such indicators are useful proxies for innovation outputs. A given increment to knowledge typically results from R&D expenditures over a number of years. This process may be modeled using the concept of an R&D "stock" which summarizes the effect of expenditures over several years. In turn, increments to knowledge by the firm over a number of years may be considered to contribute to a "stock" of technical knowledge which affects its overall productivity.

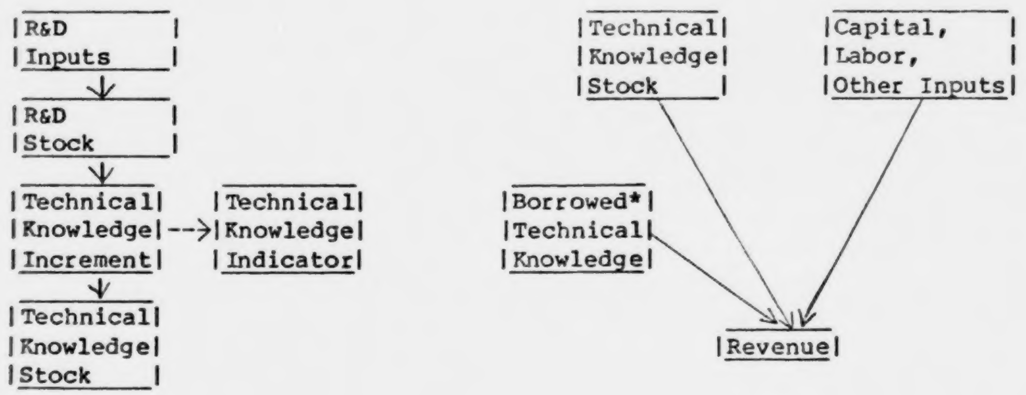
The Knowledge Production Function equation developed in this study relates measures of the R&D stock to the number of patent applications per year. Rather than focusing on the R&D expenditures that result in an individual patent, this analysis develops an equation that describes the average relationship between a firm's R&D inputs and its innovation outputs.

---

* See References at the end of the report.

Figure 2-1

The Research and Production Process



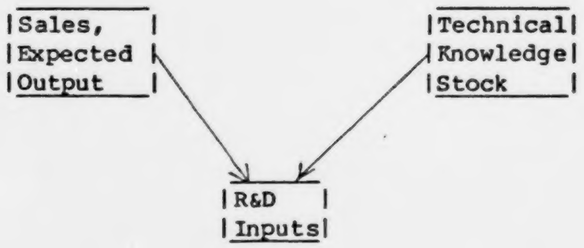
Knowledge Production Function

R&D Inputs
R&D Stock
Technical Knowledge Increment
Technical Knowledge Indicator
Technical Knowledge Stock

Revenue Function

Technical Knowledge Stock
Capital, Labor, Other Inputs
Borrowed* Technical Knowledge
Revenue

R&D Demand Equation



Sales, Expected Output
Technical Knowledge Stock
R&D Inputs

---

* Not represented in our equations

Given this structural relationship, TSCA impacts may be represented as an adjustment factor that reduces the average productivity of the R&D stock or reduces the probability that a given innovation will be successfully achieved. Increments to knowledge may relate to new or improved products, or to processes for manufacturing existing products more cheaply. Since TSCA may affect product and process of innovations differently, one version of the KPF explicitly recognizes the product/process distinction.

Revenue Function*

The right side of Figure 2-1 shows the firm's revenue as a function of capital, labor, other material inputs, and the stock of technical knowledge described above. The Revenue Function is derived from two distinct underlying structural relationships. The price a firm receives for its product is given by an inverse demand function, which relates the price to an index of product quality and to the quantity produced and sold. The firm can sell a given quantity at a higher price if it improves the product's quality, often in the form of a new product. This is precisely the effect of product innovation.

The other relationship underlying the revenue function is the production function, which relates the quantity produced to the productive inputs. When the firm's production efficiency is improved, it can produce more output with a given level of inputs. Such an increase in production efficiency can be brought about by process innovation.

If it were possible to observe separately the price of the firm's output and the physical quantity it sells, then these two underlying relationships could be estimated separately. Increase in the stock of product knowledge would show up in the inverse demand function as a higher price at any given output level. Increases in the stock of process knowledge would show up as greater physical output produced by any given level of conventional inputs. Thus, there would be a very natural way of distinguishing product and process innovations.

---

* This is not the Revenue Function of Duality Theory, in which revenues are a function of input and output prices.

Unfortunately, firms do not report their physical production nor sales price; they report only the product of the two -- revenues. Therefore, the effect of increases in the firm's knowledge stock on revenues is observed in terms of the combined effects of product innovation (which increases price) and process innovation (which increases quantity), while holding constant the quantities of other inputs. However, it is possible to develop distinct measures of the product and process knowledge stocks and examine the effects of each on revenues. This approach is developed in more detail in Chapter 4.

The impacts of TSCA on output can be represented if it is assumed that TSCA leads to a change in the overall level of the knowledge stock or a change in the marginal productivity of the knowledge stock.

R&D Demand Equation

The R&D Demand Equation is the final element of the innovation model and closes the feedback loop by specifying that R&D expenditures (inputs to the Knowledge Production Function) depend on the firm's expectations of future revenues growth which, in turn, results from increases in the stock of knowledge via the two other equations. This is shown at the bottom of Figure 2-1.

Our formulation of the R&D Demand Equation is based on the assumption that the firm will invest in R&D up to the point at which the added present value of the benefits equals the added cost. Innovation is assumed to result in reductions in unit production costs, or the equivalent effect of new products. Benefits of these cost savings are assumed proportional to total sales. The R&D Demand Equation explicitly incorporates the specification of the Knowledge Production Function, with two forms of the R&D Demand Equation derived. In the first, R&D expenditures are given as a function of last year's R&D expenditures, current sales, and the anticipated growth rate of revenues. In the second, the intensity of R&D -- R&D as a percentage of sales -- is a function of last year's R&D intensity and the expected growth rate of revenues.

The specifications of the R&D Demand Equation include parameters for the firm's discount rate, the rate of depreciation of technical knowledge, and the gestation lag between innovation and benefits for a given level of quality. The coefficients obtained from the empirical estimation can be tested to determine whether they imply reasonable values for these parameters. The model has a number of simplifying assumptions. Firms are assumed to forecast correctly (except for a random error term) the growth rate of future revenues and the benefits accruing to them in all future periods resulting from current R&D efforts. Cost savings are assumed to be a constant fraction of price in each period.

Impacts of TSCA

The most likely impact of TSCA is to change the parameters of the Knowledge Production Function (KPF). These parameters capture the cost of innovation. Assuming that expenditures for pre-market testing would be included in R&D budgets, then R&D expenditures per product innovation are expected to rise. Depending on how it is applied, TSCA could also change the parameters of the Revenue Equation. For example, EPA has the right to restrict the uses of a new product. This could reduce the expected increase in sales per product innovation.

To analyze the effects of TSCA using these three equation requires additional information from other sources. With a reasonable estimate of the reduction in the rate of product innovation that TSCA will cause, the Revenue Equation results can be used to estimate the resulting effect on productivity, ignoring TSCA's potential effects on the usefulness of innovations once produced. Alternatively, one could assume that total R&D spending will not be affected. Then, with an estimate of the cost of the testing and other requirements (perhaps on a percentage basis), the KPF can be used to estimate how this reduction in "effective" R&D would affect innovation, and the Revenue Equation would translate this into a productivity effect (again ignoring effects on the usefulness of new knowledge).

However, it is expected that R&D spending will be affected by TSCA. It would be useful, therefore, to predict the effect of TSCA on R&D spending. This is why an R&D Demand Equation is established. Such an equation captures the dependence of the level of R&D on the cost of knowledge (as captured by the KPF) and the benefits of knowledge (as captured by the Revenue Equation). If the effects of TSCA on these costs and benefits can be estimated, then it is possible to predict the effect on R&D, and hence on innovation and productivity.

Definitions and Sources of Data

Two main types of data are used in the analysis: 1) financial data, and 2) measures of innovation, in this case, counts of patent applications. Financial data for firms, including sales revenue, R&D expenditures, capital expenditures and employment, were taken from firms' 10-K and annual reports. In addition, line of business financial data was obtained from the U.S. Federal Trade Commission (FTC). The FTC Line of Business data base disaggregates firm-level financial data into the various lines of business in which the firm is active. Value-added data by four digit SIC group were obtained from the U.S. Census of Manufactures. Counts of chemical patent applications are from the U.S. Patent Office, Office of Technology Assessment and Forecasting (OTAF). Two patent data bases were developed; one that includes patent applications by firms in the firm-level financial data base and one that includes patent applications by firms in the FTC Line of Business data base.

Aggregate Firm-Level Data

The selection of firms for the sample was based on the firm's involvement in the chemical industry and data availability . For this sample of chemical firms, financial data were collected from their 10-K and annual reports, for the period 1960-1981. However, not all firms have observations for all years. All financial variables are deflated by appropriate price indices. The deflators used are:

- Sales - wholesale price index for chemicals and allied products (SIC 28),
- Capital Expenditures - *Chemical Week* plant cost index,
- R&D Expenditures - Biomedical R&D index.

A data base containing patent applications filed by firms in the firm-level financial data base was compiled. The patent data include patent applications for patents that were granted by the U.S. Patent Office between 1969 and 1980. In other words, applications filed during the period analyzed but granted later are not included. This definition causes observations before about 1968 and after 1978 to be incomplete; so the analysis was done only for data falling within this 1968 to 1978 time period. The equations that distinguish between product and process patents are estimated using the firm-level data sets, with a few observations dropped for reasons discussed in Chapters 3, 4, and 5. The major difference in the data is that patents have been classified as product or process on the basis of the patent titles provided by the U.S. Patent Office.*

The selection procedure biased the sample to large diversified firms, whose innovative behavior may differ from that of smaller, more specialized firms. The basic data set used to estimate the Knowledge Production Function (presented in Chapter 3) consists of 436 observations for 63 firms that have positive values for both patent applications and R&D expenditures. (Some additional work was done including firms with zero or missing values for these variables.) The distribution of the 63 firms by firm size is shown in Table 2-1.

The basic data set used to estimate the Revenue Function (presented in Chapter 4) consists of 782 observations for 100 firms that have positive values for inputs and sales variables and that have associated 4-digit SIC codes. (As in the case of the KPF, some additional work was done including

---

* Appendix 2-B presents the results of a test of the reliability of the classification procedures.

Table 2-1.

Distribution of Firms in the Firm-Level Financial Data Bases by Size of Firm

Size of Firm Sales (million $)	Knowledge Production Function		Revenue Function		R&D Demand Function	
	Number of Firms	Percent of Total	Number of Firms	Percent of Total	Number of Firms	Percent of Total
Less than 10	0	0	7	7	6	7
10-99	8	13	26	26	19	22
100-999	26	41	36	36	30	35
1,000-2,500	18	29	19	19	19	22
Over 2,500	11	17	12	12	11	13
Total	63		100		85	

firms with zero or missing values for R&D inputs.) The distribution of the 100 firms by firm size is also shown in Table 2-1. The revenue function also includes employment, from the annual reports, and a time trend to measure changes in total factor productivity. Although material expenditures are also an important input, no firm-level data on them were available.

The basic data set for the R&D Demand Equation (present in Chapter 5) was smaller than that for the Revenue Function. Each firm had to have at least six consecutive years of sales data and non-zero values for current and lagged R&D expenditures, because the growth rate term used in the equations is a four year average and because both current and lagged R&D appear in the equations. This yielded a subset of 85 firms out of the larger data base of 100, with a total of 524 observations. The distribution of the 85 firms by sales size is shown in Table 2-1.

Characteristics of firms included in the analysis are presented in Appendix 2-A, and are briefly summarized here. As discussed above, the sample is biased towards large firms in terms of sales. These firms also tend to be diversified, with, on average three 4-digit SIC codes assigned to each.

Total R&D expenditures in the sample average 2.2 percent of total sales. Small firms, with sales of less than $10 million, have an R&D to sales ratio that is much higher than that of the other sales groups during the 1973 to 1980 period. All sales groups, except the less than $10 million, show roughly the same trend over time: a steady decline in the ratio of R&D to sales over the period 1960 to 1974 and a relatively stable relationship after that with a trough in 1978 and 1979 which corresponds to the economic recession.

The total number of patents in the sample for the period 1968 to 1978 is 24,705, with 85 percent belonging to firms with sales greater than $1,000 million. In terms of the average ratio of chemical patents to R&D expenditures, there is a slight downward trend over the time period for the largest firms, and a rise followed by a decline for the middle groups, $100-999 million sales. Trends for the smallest sales group are more difficult to detect.

Line of Business Data

The Line of Business (LB) financial data base is described in detail in FTC publications*, and is briefly summarized here. The data consist of financial, operating and asset information for about 440 large U.S. manufacturing corporations for the period 1974-1976, with data for 1977 now becoming available. The exact sample varies slightly from year to year. For reasons described below, the patent/R&D file used in the line of business analysis in this study includes only data for 1974. Each firm's data are broken down into separate lines of business (LB) units which correspond to 4-digit SIC codes.** The LB data covers the domestic, unregulated part of firms' operations. Foreign operations and domestic regulated operations are available only as overall totals.

The two items of information from the FTC Line of Business data used in this study are R&D expenditures and LB sales. LB sales are defined as: "total net operating revenues and transfers." R&D expenditures were classified as basic or applied and as company-financed, federally-financed, or other outsider-financed. Applied R&D was reported at the LB level while basic R&D was reported only at the firm level. There were some anomalies in firm reporting behavior; for example, many reported zero basic R&D while others reported all R&D as basic even though this is unlikely for many firms of this size. As part of an unrelated study of inter-industry technology flows, F.M. Scherer developed a data base that links patent and Line of Business R&D data, for 1974 LB data. This linked patent/R&D data base was used for the LB section of this analysis. In developing his patent/R&D data base, Scherer allocated basic R&D to the individual LB's in proportion to private applied R&D outlays in each LB.* Also, certain companies allocated some R&D to foreign or domestic regulated activities which are not broken down to the LB level. These amounts were not included in the LB R&D totals. Non-company financed R&D represents a very small share of total R&D in chemical industry LB's, almost always less than one percent, and hence can be safely ignored.

---

* Federal Trade Commission (1982).

** Some data items, such as basic R&D, were reported only for the firms as a whole. As described below, basic R&D was allocated to individual LB's.

In developing his linked patent/R&D data base, Scherer assumed a lag of nine months between R&D expenditures and patent applications and a further delay of 19 months for the patent to be granted.** Thus the total time lag between innovation and granting of patent is estimated to be 28 months. Based on this time lag, he set the time span for his sample of patents to be the ten-month period of June 1976 through March 1977. The mid-point of this period is 28 months after the mid-point of the R&D data he wanted to use, i.e. June 30, 1974. Scherer identified all patents attributable to the 443 firms in the 1974 LB data base that were granted during that period

---

* Scherer (1984). This was done only for the 1974 LB data.

** Scherer (1984).

(15,112), and matched the patents to individual LB units.* In some cases, fractions of patents were attributed to multiple LB units. There is an inconsistency between the LB patent data, where patents were chosen on the basis of date granted, and the firm-level patent data where patents were classified by application date. However, this difference is not likely to be a large problem, because a zero year lag was assumed in the firm-level regressions, which is not inconsistent with the nine month lag for applications assumed here. Therefore, the main effect of using patents granted is the additional noise introduced into the patent-R&D relationship by variations in the time taken to grant a patent.

Each patent was classified in terms of the nature of the innovation (e.g., product or process), the industries of origin and use, the degree of complexity, and other characteristics. Finally, since the sampling period covered only 10 months, patent counts of each LB were inflated by the factor 1.2009 to provide counts consistent with annual R&D expenditures.

The sample used in this part of the study includes all LB's classified as non-pharmaceutical chemicals (i.e., all LB 2800 codes except 2807 and 2808) plus petroleum refining (LB code 2901). This yielded a total of 399 LB units belonging to 135 firms; 257 of these LB's have non-zero observations for both patents and R&D. Only 30 of the firms in this sample are also included in the firm-level data sample. The rest of the LB's belong to firms not classified overall as being in the chemical industry.

---

* The details of the procedure and the problems involved are given in Scherer (1984) and are too lengthy to be repeated here.

Table 2-2 shows the total number of patents on an annual basis by LB code and LB sales group. Overall, the sample has 4,180 patents. Over one-half were generated by LB units with between $100 million and $1,000 million in sales, while well over one-half the patents were produced by LB units in LB 2804 (SIC 2821, plastic materials and resins), LB 2813 (SIC 2865,69 industrial organic chemicals, not elsewhere classified, except gum and wood chemicals), and LB 2901 (SIC 2911, petroleum refining). The distribution of LB units is not shown for reasons of confidentiality.

Table 2-3 shows the average number of patents per million dollars of R&D expenditures by LB code and sales group. Most of the ratios lie between one and 10 patents. Observations with very high ratios tend to be cases with fractional patents and very low R&D expenditures, and hence may be regarded as artifacts of the data construction process. There does appear to be a negative relationship between LB size and patents per million dollars, similar to the relationship in the firm-level data.

Principal Econometric Problems

The principle econometric problems encountered in this work are "errors in variables," because patents are an imperfect measure of innovation, and "fixed effects," because firms differ in unobserved ways that are potentially correlated with the variables of interest. These two problems interact, because solutions for one tend to aggravate the other.

The basic dilemma may be stated simply. If patents turn out to be a "noisy" measure of innovation, then patent applications in any one year are an imperfect measure of innovation. Therefore, the use of patents to measure knowledge leads to biased coefficients. If, however, the "measurement

Table 2-2.

Total Patents by LB Code and Sales Size Group

LB Sales Group (Million $)

LB Code	Under 10	10-99	100-999	1000-2499	Over 2500
2801	0	0	33	0	0
2802	0	6	--	0	0
2803	--	169	186	0	0
2804	0	170	648	0	0
2805	0	40	--	0	0
2806	0	11	144	--	0
2809	0	13	63	0	0
2810	0	21	75	--	0
2811	0	15	97	0	0
2851*	0	281	476	299	0
2814	0	5	18	0	0
2815	0	78	216	0	0
2816	--	5	0	0	0
2817	25	155	47	0	0
2901	0	0	0	50	638

* For reasons of confidentiality, results in LB group 2812 and 2813 have been aggregated (2851) and some other results have been omitted.

Table 2-3.

Ratio of Patents to Million Dollars of
R&D by LB Code and Sales Group

LB Sales Group (Million $)

LB Code	Under 10	10-99	100-999	1000-2499	Over 2500
2801	0	0	4.54	0	0
2802	0	2.33	--	0	0
2803	--	6.98	7.04	0	0
2804	0	8.55	5.02	0	0
2805	0	4.55	--	0	0
2806	0	1.24	3.01	--	0
2809	0	0.61	1.63	0	0
2810	0	2.47	3.60	--	0
2811	0	2.70	1.90	0	0
2851*	0	7.56	3.82	2.63	0
2814	0	1.44	35.96	0	0
2815	0	9.76	4.53	0	0
2816	--	1.35	0	0	0
2817	125.36	11.51	6.42	0	0
2901	0	0	0	8.00	13.82

* For reasons of confidentiality, the results for LB Group 2851 represent a weighted average of the results for LB Groups 2812 and 2813 and some other results have also been omitted.

error" is purely random (a not unreasonable assumption), then the average over many years of the patents granted to a firm will be a better measure of innovation, because the noise will tend to average out. Thus the measurement error problem implies that it is better to estimate the relationships on the basis of the variance "between" firm averages, since "within" a firm over time there is mostly noise.

On the other hand, it is often assumed that the error structure in the types of models estimated in this study are of the form:

$$Y_{it} = B\,X_{it} + h_i + e_{it} \qquad (2\text{-}1)$$

where the h_i are unobserved characteristics of the firms such as management skill or the propensity to patent. If we simply regress Y on X, or if we take firm averages as implied above and use them in the regression, we get unbiased estimates of B only if h_i is uncorrelated with the X's. Often it is not plausible to assume such a lack of correlation. In particular, if one of the X's is patents and h_i is the firm's propensity to patent, they will always be correlated. The usual prescription is to estimate the h_i's by using firm dummies, or equivalently, to get rid of them by running the regression on the deviations from firm means. This is referred to as the "within" estimation. Unfortunately, this "within" regression aggravates, as a practical matter, the effect of the errors in variables problem discussed above, creating a dilemma that has no attractive resolution.

The textbook solution is to use instrumental variables and the within regression. This was tried and it did not work particularly well, probably for two reasons. First, there are no good instruments. Second, there does not appear to be much within firm variance in the variables of interest that is not measurement error. This is not surprising based on theory, and is often observed in this type of data. See, for example, Griliches and Mairesse (1981).

In this data set, the fact is that "within" regressions do not work for many versions of our model equations. In particular, to estimate equations that make a distinction between product and process innovation, the "between" results must be used. This necessitates some consideration of the seriousness and likely direction of the potential fixed effects bias; this will be discussed in the context of the individual equations in the following chapters.

APPENDIX 2-A.
Characteristics of Firms.

This appendix describes some of the characteristics of chemical firms based on the financial and patent data base developed for the TSCA study. It consists mainly of tables and graphs, with some results of analysis of variance presented as well. The data are analyzed according to the following dimensions:

- time: by year;
- firm sales: less than $10 million, $10-99 million, $100-999 million, $1000-2499 million, $2500 million and over;
- 3 digit major SIC code for firm (Disclosure Data Base); and
- 4 digit SIC codes including all major activities of firms (from Dun & Bradstreet).

The variables examined are counts of firms, firm sales, R&D expenditures, total chemical patents, the share of product in total patents, and the relationships of these variables, such as the ratio of chemical patents to R&D expenditures. Table 2A-1 shows a summary of this data for each firm in the data base. Table 2A-2 gives the definitions of the SIC codes examined in the study.

Distribution of Firms

Table 2A-3 shows the distribution of firms in the financial data base by firm sales size (based on 1979 or the most recently available data) and major 3-digit SIC code.* Naturally, because of the difficulty of obtaining financial data for small companies, the sample is biased toward large firms.

---

* "9999" indicates all other SIC codes.

TABLE 2A-1.
SUMMARY OF FIRM DATA

	Company	Firm Sales	R&D Expenditures	Number of Patents	Ratio of Product Patents to Total Patents
1	ACETO	79.4	0.242	.	.
2	AIRPROD	1230.0	23.000	296	0.47973
3	AKZONA	1012.0	25.300	315	0.64127
4	ALBERTO	190.0	2.800	22	0.86364
5	ALCOLAC	34.2	0.928	16	0.81250
6	ALLIED	4332.0	70.000	1560	0.52564
7	ALOE	3.5	0.010	.	.
8	AMERCYAN	3187.0	124.000	2485	0.66600
9	ANSUL	110.9	1.429	.	.
10	ATI	28.1	0.200	4	1.00000
11	AVON	2378.0	24.900	27	0.62963
12	BALCHEM	3.6	0.046	6	0.50000
13	BETZ	.	.	144	0.69444
14	BRISTOL	2753.0	103.000	366	0.79508
15	CABOT	1122.0	17.300	123	0.56098
16	CARBOL	55.6	1.471	.	.
17	CELANESE	3146.0	92.000	864	0.37731
18	CHATTEM	47.3	0.402	.	.
19	CHEMET	547.2	10.045	57	0.24561
20	CHESEBR	1174.0	9.700	4	0.75000
21	CHOMER	10.7	0.700	4	1.00000
22	CHURCH	97.6	2.830	8	0.50000
23	COLGATE	4494.0	41.100	725	0.79862
24	COLUMBIA	.	.	1	0.00000
25	CROMPTON	240.0	4.000	28	0.57143
26	DART	2403.0	14.200	207	0.52174
27	DEL	44.7	0.610	4	0.75000
28	DESOTO	326.5	13.500	138	0.73913
29	DETREX	75.6	1.050	15	0.40000
30	DEXTER	457.0	13.600	34	0.85294
31	DIAMOND	2394.0	42.000	411	0.53285
32	DOW	9255.0	269.000	3879	0.53364
33	DOWCORN	610.2	40.200	594	0.72054
34	DUPONT	12572.0	415.000	5016	0.60766
35	EARTH	406.0	0.627	.	.
36	ECONLAB	464.4	4.100	42	0.47619
37	ESSEX	127.0	2.200	2	0.50000
38	ETHYL	1657.0	[illegible]	[illegible]	0.44444
39	FABERGE	245.0	1.000	4	0.75000
40	FAIRMNT	9.0	0.150	1	1.00000
41	FERRO	592.0	2.260	100	0.63000
42	FILTROL	101.7	1.395	24	0.79167
43	FMC	3308.0	86.300	.	.
44	FREEPRT	487.7	41.300	39	0.20513
45	GLASRCI	11.9	0.137	15	0.33333
46	GROW	214.0	4.900	16	0.56250
47	GRTLAKES	125.0	3.300	17	0.52941
48	GUARDC	51.0	2.200	.	.
49	HELECRT	131.0	1.600	6	0.33333
50	HERCULES	2345.0	47.000	648	0.63272
51	HOOKER	2086.0	28.000	824	0.52670
52	HUNT	166.0	5.400	19	0.57895
53	IMC	1475.0	6.600	100	0.51000
54	IMPERIO	[illegible]	37.000	.	.
55	INTLFF	409.0	27.000	467	0.70878
56	INSILCO	[illegible]	[illegible]	.	.

TABLE 2A-1. (continued)

	Company	Firm Sales	R&D Expenditures	Number of Patents	Ratio of Product Patents to Total Patents
57	LAWTER	68.20	1.110	4	1.00000
58	LEARONAL	114.00	0.500	13	0.30767
59	LOCTITE	161.00	3.900	32	0.71875
60	LUBRIZOL	725.00	32.500	274	0.82482
61	MALLIN	392.50	12.750	.	.
62	MEDAL	80.85	0.574	.	.
63	MERLNOR	55.69	0.628	.	.
64	MONSANTO	6193.00	174.000	2984	0.52815
65	MORTON	732.00	19.200	247	0.87952
66	NALCO	579.00	21.600	493	0.43611
67	NENUKE	65.80	5.200	16	0.68750
68	NESTLE	10.70	0.079	1	0.00000
69	NLINDUST	1810.00	27.000	249	0.51004
70	NOXELL	179.70	2.000	1	1.00000
71	OAKITE	60.20	1.100	11	0.63636
72	OLIN	1778.00	26.000	844	0.50237
73	ORROX	7.70	0.180	.	.
74	PARK	17.10	0.000	11	0.72727
75	PEARSAL	14.98	0.487	1	0.00000
76	PENWALT	1079.00	24.000	480	0.64583
77	PETROL	186.90	5.600	.	.
78	PLASMINE	1.20	0.008	4	1.00000
79	POLYRES	1.01	0.410	6	0.16667
80	PPG	3092.00	83.300	.	.
81	PRATT	97.50	1.400	12	0.83333
82	PRODRES	41.50	1.200	7	1.00000
83	PUREX	628.70	2.700	.	.
84	QUAKER	92.10	3.213	.	.
85	REALEX	25.80	0.340	.	.
86	REDKEN	65.00	1.820	5	0.60000
87	REICH	1095.30	6.445	90	0.72222
88	RELIUNI	189.60	0.928	.	.
89	REVLON	1718.00	40.500	17	0.64706
90	ROHM	1590.00	54.000	757	0.71466
91	RPM	101.00	0.379	.	.
92	SHERWIN	1196.00	7.100	140	0.60000
93	SPECIAL	5.31	0.226	.	.
94	STAUFFER	1526.00	37.000	1252	0.72364
95	STEPAN	145.00	4.300	26	0.69231
96	SUN	468.70	9.200	66	0.72727
97	TEXASG	789.00	2.400	18	0.00000
98	THIOKOL	627.20	8.600	.	.
99	UNICHEM	6.31	0.079	.	.
100	UNIONCAR	9177.00	161.000	2200	0.58409
101	UNIVAR	713.10	1.242	.	.
102	VALSPAR	121.00	2.700	7	0.42857
103	VIRCHEM	131.30	2.088	16	0.00000
104	WESTCHEM	60.50	0.650	1	1.00000
105	WHEELAB	947.00	8.400	13	0.00000

TABLE 2A-2.
INDUSTRIAL SECTORS, BY SIC NUMBER, INCLUDED IN STUDY

- 28 Chemicals & Allied Products
  - 281 Industrial Inorganic Chemicals
    - 2812 Alkalies & Chlorine
    - 2813 Industrial Gases
    - 2816 Inorganic Pigments
    - 2819 Industrial Inorganics NEC
  - 282 Plastic Materials & Synthetic Resins, Rubber, Other Fibers, Except Glass
    - 2821 Plastic Materials, Synthetic Resins & Non-Vulcanized Elastomers
    - 2822 Synthetic Rubber
    - 2823 Cellulosic Man-Made Fibers
    - 2824 Synthetic Organic Fibers, except Cellulosic
  - 284 Soaps, Detergents, Cleaning Preparations, Perfumes, Cosmetics, Toilet Preparations
    - 2841 Soap & Other Detergents, Except Specialty Cleaners
    - 2842 Specialty Cleaning, Polishing & Sanitation Preparations
    - 2843 Surface Active Agents, Finishing Agents, Sulfonated Oils & Assistants
    - 2844 Perfumes, Cosmetics & Other Tiolet Preparations
  - 285 Paints, Varnishes, Lacquers, Enamels & Allied Products
    - 2851 (Same description as 285)
  - 286 Industrial Organic Chemicals
    - 2861 Gum & Wood Chemicals
    - 2865 Cyclic (Coal Tar) Crudes, Cyclic Intermediates, Dyes & Organic Pigments
    - 2869 Industrial Organics NEC
  - 287 Agricultural Chemicals
    - 2873 Nitrogenous Fertilizers
    - 2874 Phosphatic Fertilizers
    - 2875 Fertilizers, Mixing Only
  - 289 Miscellaneous Chemical Products
    - 2891 Adhesives & Sealants
    - 2892 Explosives
    - 2893 Printing Ink
    - 2895 Carbon Black
    - 2899 Chemicals & Chemical Products NEC
- 29 Petroleum Refining & Related Industries
  - 291 Petroleum Refining
    - 2911 Petroleum Refining

TABLE 2A-3.
COMPANY FREQUENCY BY MAJOR SIC CODE AND SALES GROUP (1979)

Major SIC

Salesgroup

FREQUENCY PERCENT	.	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2810	1	2	4	5	12	7	30
	.	1.96	3.92	4.90	11.76	6.86	29.41
2820	0	0	1	1	0	0	2
	.	0.00	0.98	0.98	0.00	0.00	1.96
2830	0	0	1	1	0	0	2
	.	0.00	0.98	0.98	0.00	0.00	1.96
2840	0	2	10	8	3	2	25
	.	1.96	9.80	7.84	2.94	1.96	24.51
2850	0	0	3	6	1	0	10
	.	0.00	2.94	5.88	0.98	0.00	9.80
2860	0	0	0	2	0	0	2
	.	0.00	0.00	1.96	0.00	0.00	1.96
2870	0	0	1	0	1	0	2
	.	0.00	0.98	0.00	0.98	0.00	1.96
2890	2	4	3	8	1	0	16
	.	3.92	2.94	7.84	0.98	0.00	15.69
2910	0	0	0	1	0	1	2
	.	0.00	0.00	0.98	0.00	0.98	1.96
9999	0	0	3	5	1	2	11
	.	0.00	2.94	4.90	0.98	1.96	10.78
TOTAL	.	8	26	37	19	12	102
	.	7.84	25.49	36.27	18.63	11.76	100.00

Of the 102 firms in the sample, 68 have sales greater than $100 million, while only eight have sales less than $10 million. Thirty of the firms are found in SIC group 281, Industrial Chemicals, and another 25 are found in SIC group 284, Soaps, Detergents, Cleansing Preparations, Cosmetics, and Toilet Preparations. The firms in group 281 tend to be quite large (and diversified), with 19 having sales over $1 billion, while the average size of the firms in SIC 284 is significantly smaller, the mode cell being $10-99 million.

Table 2A-4 shows the distribution of firm activities in the 4 digit SIC code areas. The total number of entries is 298, indicating an average of three 4-digit activities per firm. Comparing the column totals with those in Table 2A-3, the average number of activities of firms increases with size, ranging from 1.0 for firms under $10 million to 6.7 for firms over $2500 million.*

Combining the four-digit codes into their respective 3-digit codes yields the following counts for Table 2A-4:**

SIC	No. of Firms
281	66
282	45
283	22
284	62
285	14
286	40
287	20
289	42
291	2

---

* Non-chemical activities are not fully included in these counts.

** These have not been corrected for possible double-counting.

TABLE 2A-4.

Company Frequency by SIC Code and Salesgroup

SIC FREQUENCY PERCENT	.	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2812	0	0	3	0	9	4	16
	.	0.00	1.01	0.00	3.02	1.34	5.37
2813	0	0	0	1	1	1	3
	.	0.00	0.00	0.34	0.34	0.34	1.01
2816	0	1	1	1	5	3	11
	.	0.34	0.34	0.34	1.68	1.01	3.69
2818	0	0	0	0	2	2	4
	.	0.00	0.00	0.00	0.67	0.67	1.34
2819	1	0	5	8	11	8	32
	.	0.00	1.68	2.68	3.69	2.68	10.74
2821	0	0	6	6	11	9	32
	.	0.00	2.01	2.01	3.69	3.02	10.74
2822	0	0	0	1	1	1	3
	.	0.00	0.00	0.34	0.34	0.34	1.01
2823	0	0	0	0	2	3	5
	.	0.00	0.00	0.00	0.67	1.01	1.68
2824	0	0	0	0	0	5	5
	.	0.00	0.00	0.00	0.00	1.68	1.68
2831	0	0	0	0	0	1	1
	.	0.00	0.00	0.00	0.00	0.34	0.34
2833	0	0	2	1	1	2	6
	.	0.00	0.67	0.34	0.34	0.67	2.01
TOTAL	.	7	59	73	86	73	298
	.	2.35	19.80	24.50	28.86	24.50	100.00

TABLE 2A-4. (continued) Company Frequency by SIC Code and Salesgroup

SIC FREQUENCY PERCENT	Salesgroup .	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2834	0	0	6	2	4	3	15
	.	0.00	2.01	0.67	1.34	1.01	5.03
2841	0	0	4	5	0	2	11
	.	0.00	1.34	1.68	0.00	0.67	3.69
2842	0	0	8	7	0	1	16
	.	0.00	2.68	2.35	0.00	0.34	5.37
2843	0	0	0	1	0	0	1
	.	0.00	0.00	0.34	0.00	0.00	0.34
2844	0	1	6	6	3	3	19
	.	0.34	2.01	2.01	1.01	1.01	6.38
2851	0	0	4	7	1	2	14
	.	0.00	1.34	2.35	0.34	0.67	4.70
2861	0	0	0	1	2	2	5
	.	0.00	0.00	0.34	0.67	0.67	1.68
2865	0	1	0	1	2	3	7
	.	0.34	0.00	0.34	0.67	1.01	2.35
2869	0	1	4	6	10	7	28
	.	0.34	1.34	2.01	3.36	2.35	9.40
2870	0	0	0	0	1	0	1
	.	0.00	0.00	0.00	0.34	0.00	0.34
2871	0	0	0	0	0	1	1
	.	0.00	0.00	0.00	0.00	0.34	0.34
TOTAL	.	7	59	73	86	73	298
	.	2.35	19.80	24.50	28.86	24.50	100.00

TABLE 2A-4. (continued) Company Frequency by SIC Code and Salesgroup

SIC

Salesgroup

FREQUENCY PERCENT	.	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2872	0 .	0 0.00	0 0.00	0 0.00	0 0.00	1 0.34	1 0.34
2873	0 .	0 0.00	0 0.00	0 0.00	3 1.01	2 0.67	5 1.68
2874	0 .	0 0.00	0 0.00	0 0.00	1 0.34	0 0.00	1 0.34
2879	0 .	1 0.34	1 0.34	2 0.67	4 1.34	3 1.01	11 3.69
2891	0 .	0 0.00	3 1.01	4 1.34	3 1.01	0 0.00	10 3.36
2892	0 .	0 0.00	0 0.00	0 0.00	2 0.67	2 0.67	4 1.34
2893	0 .	0 0.00	2 0.67	2 0.67	0 0.00	0 0.00	4 1.34
2895	0 .	0 0.00	0 0.00	0 0.00	1 0.34	0 0.00	1 0.34
2899	0 .	2 0.67	4 1.34	10 3.36	5 1.68	2 0.67	23 7.72
2911	0 .	0 0.00	0 0.00	1 0.34	1 0.34	0 0.00	2 0.67
TOTAL	. .	7 2.35	59 19.80	73 24.50	86 28.86	73 24.50	298 100.00

Comparing these results with Table 2A-3 shows that the SIC codes which appeared to be very skimpily represented in the data base, such as 282, 283, 286 and 287, are much better represented when firm diversification is taken into account. The problem is to coax the firm-level data to yield results about division-level activities.

Table 2A-5 shows the equivalent distribution in Table 2A-3 for the product/process patent data base. Of the 80 firms in the patent data base, 77 also have sales data. The relative proportions are about the same, with firms in SIC group 281 having a somewhat greater share (36 percent vs. 29 percent), and the distribution by size is also quite close.

Annual Sales

Table 2A-6 shows the total annual sales of all companies in each cell corresponding to Table 2A-3. Total annual sales in the sample are $114.2 billion, of which $69.0 billion (60 percent) is accounted for by firms in SIC group 281 with over $1 billion in sales. Another 14 percent of sales is contributed by firms in SIC group 284.

R&D Expenditures

Table 2A-7 shows total R&D expenditures for all firms in each cell corresponding to Table 2A-3. Total R&D expenditures in the sample are $2.49 billion, or 2.2 percent of total sales. The distribution of R&D expenditures is similar to that for sales for firms with sales less than $100 million, but firms with sales of $1000-2499 million have slightly lower shares of R&D

TABLE 2A-5.

Company Frequency by Major SIC Code and Salesgroup

Major SIC FREQUENCY PERCENT	Salesgroup .	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
.	0 .	4 .	9 .	9 .	0 .	3 .	. .
2810	1 .	0 0.00	4 5.19	5 6.49	12 15.58	7 9.09	28 36.36
2820	0 .	0 0.00	1 1.30	1 1.30	0 0.00	0 0.00	2 2.60
2840	0 .	0 0.00	8 10.39	7 9.09	3 3.90	2 2.60	20 25.97
2850	0 .	0 0.00	1 1.30	3 3.90	1 1.30	0 0.00	5 6.49
2860	0 .	0 0.00	0 0.00	2 2.60	0 0.00	0 0.00	2 2.60
2870	0 .	0 0.00	0 0.00	0 0.00	1 1.30	0 0.00	1 1.30
2890	2 .	4 5.19	3 3.90	7 9.09	1 1.30	0 0.00	15 19.48
9999	0 .	0 0.00	0 0.00	3 3.90	1 1.30	0 0.00	4 5.19
TOTAL	. .	4 5.19	17 22.08	28 36.36	19 24.68	9 11.69	77 100.00

TABLE 2A-6.
1979 Sales by Major SIC Code and Salesgroup

Major SIC		Salesgroup					
FREQUENCY PERCENT	.	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2810	103	13	246	1365	20993	47862	70480
	.	0.01	0.22	1.20	18.38	41.91	61.72
2820	0	0	12	610	0	0	622
	.	0.00	0.01	0.53	0.00	0.00	0.54
2830	0	0	47	393	0	0	440
	.	0.00	0.04	0.34	0.00	0.00	0.39
2840	0	10	512	2716	5270	7247	15755
	.	0.01	0.45	2.38	4.62	6.35	13.80
2850	0	0	204	1537	1196	0	2937
	.	0.00	0.18	1.35	1.05	0.00	2.57
2860	0	0	0	965	0	0	965
	.	0.00	0.00	0.85	0.00	0.00	0.85
2870	0	0	26	0	1475	0	1501
	.	0.00	0.02	0.00	1.29	0.00	1.31
2890	0	15	69	3506	1122	0	4712
	.	0.01	0.06	3.07	0.98	0.00	4.13
2910	0	0	0	406	0	5953	6359
	.	0.00	0.00	0.36	0.00	5.21	5.57
9999	0	0	284	2723	1012	6400	10419
	.	0.00	0.25	2.38	0.89	5.60	9.12
TOTAL	.	38	1401	14220	31068	67462	114189
	.	0.03	1.23	12.45	27.21	59.08	100.00

TABLE 2A-7.

1979 R & D Expenditure by Major SIC Code and Salesgroup

Major SIC

Salesgroup

FREQUENCY PERCENT		UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
2810	1	0	10	45	358	1305	1717
	.	0.02	0.39	1.79	14.38	52.48	69.06
2820	0	0	0	40	0	0	40
	.	0.00	0.01	1.62	0.00	0.00	1.62
2830	0	0	0	13	0	0	13
	.	0.00	0.02	0.52	0.00	0.00	0.54
2840	0	0	8	38	75	144	265
	.	0.00	0.31	1.52	3.02	5.79	10.64
2850	0	0	5	24	7	0	36
	.	0.00	0.20	0.97	0.29	0.00	1.46
2860	0	0	0	27	0	0	27
	.	0.00	0.00	1.07	0.00	0.00	1.07
2870	0	0	0	0	7	0	7
	.	0.00	0.01	0.00	0.27	0.00	0.28
2890	0	1	2	65	17	0	85
	.	0.02	0.08	2.62	0.70	0.00	3.41
2910	0	0	0	1	0	37	38
	.	0.00	0.00	0.03	0.00	1.49	1.51
9999	0	0	5	59	25	170	259
	.	0.00	0.20	2.37	1.02	6.82	10.40
TOTAL	.	1	30	311	489	1656	2487
	.	0.04	1.21	12.50	19.67	66.58	100.00

relative to firms with sales greater than $2500 million. Over half of total R&D expenditures came from firms in SIC group 281 with sales over $2,500 million.

This is borne out in Table 2A-8, which shows the mean and standard deviation of the ratio of R&D expenditures to sales for firms in each cell. For firms in SIC group 281, the R&D to sales ratio for the sales size $1000-2499 million is .017 and that for the sales size over $2500 million is .027 (11 and 7 observations, respectively). However, most cells have too few observations to draw firm conclusions.

Figure 2A-1 shows the trend of the average R&D to sales ratio for each sales group over the period 1960-1981. All sales groups except "less than $10 million" show roughly the same trend: a steady decline over the period 1960-1973 and a relatively stable relationship after that with another trough in 1978-1979 which corresponds to the economic recession at that time. However, the trends may be misleading because they mask the effect of changing sample size. Many more firms began reporting R&D expenditures after 1970, especially in the smaller sales groups, and this effect must be taken into account. The firms with less than $10 million sales have an R&D to sales ratio much higher than that of the other sales groups (ranging from .06 to .09 as opposed .02 to .04 for the other groups) over the period 1973-1980.* This may be due to sampling bias, however.

---

* Observations for this sales group outside this time period are based on only a very few observations.

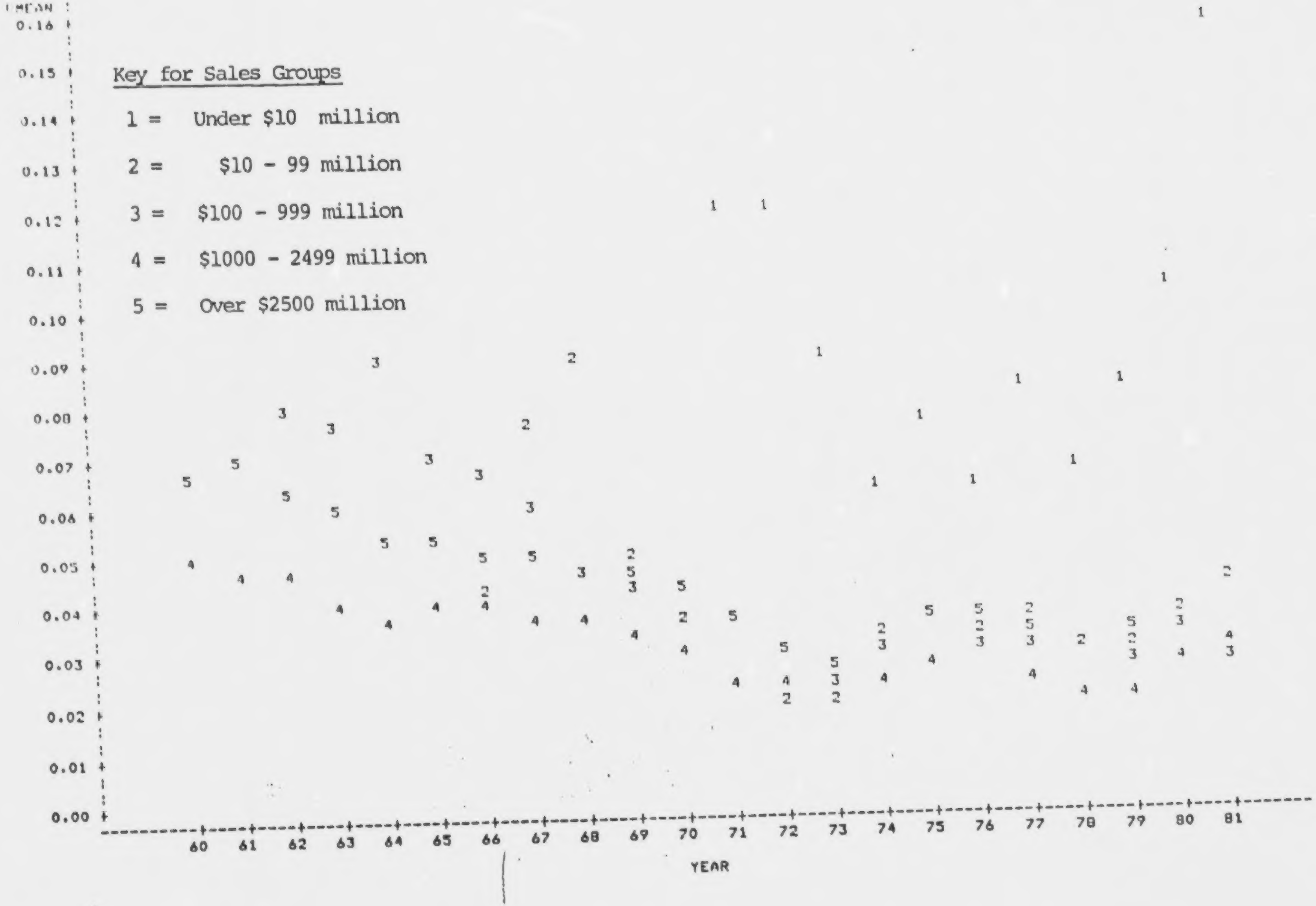
TABLE 2A-8.
Mean & Standard Deviation of 1979 R&D to Sales
Ratio by Major SIC Code and Sales Group

Major SIC Code	Sales Group Missing	Under 10	10-99	100-999	1000-2499	Over 2500
2810	0.014 (NA)	0.033 (0.014)	0.039 (0.027)	0.029 (0.021)	0.017 (0.008)	0.027 (0.008)
2820	---	---	0.012 (NA)	0.066 (NA)	---	---
2830	---	---	0.009 (NA)	0.033 (NA)	---	---
2840	---	0.008 (0.007)	0.015 (0.008)	0.014 (0.009)	0.014 (0.008)	0.023 (0.020)
2850	---	---	0.028 (0.014)	0.016 (0.015)	0.006 (NA)	---
2860	---	---	---	0.024 (0.010)	---	---
2870	---	---	0.013 (NA)	---	0.004 (NA)	
2890	---	0.111 (0.197)	0.031 (0.033)	0.020 (0.013)	0.015 (NA)	---
2910	---	---	---	0.002 (NA)	---	0.006 (NA)
9999	---	---	0.017 (0.016)	0.031 (0.036)	0.025 (NA)	0.027 (0.001)

NA: Not applicable (i.e. only one observation)

--- Not available
( ) Standard deviation

FIGURE 2A-1. Trends in Ratio of R&D to Sales



Total Chemical Patents

Data on chemical patent applications made between 1960 and 1980 were obtained for 80 chemical firms from the Office of Technology Assessment and Forecasting (OTAF) of the U.S. patent office. The data set includes all patents which were granted in the period 1968 to 1980. Because of this peculiarity of the definition, applications for years prior to 1968 and after 1977 are incomplete because many of them were not granted in the period 1968 to 1980. This needs to be remembered when interpreting the data. In some cases, results are presented only for the interval 1968 to 1978.

Table 2A-9 shows total chemical patents for each sales group and the sample total for the period 1968 to 1978. The total number of patents over the period is 24,705, with 22,207 of them (85 percent) being supplied by firms with sales of greater than $1000 million. The four firms with sales of less than $10 million produced only 127 patents over the period, while the 17 firms with $10-99 million in sales produced ony 109 patents. Among the largest firms with sales over $2500 million, the total annual number of patents declines by 37 percent over the period, but this trend is much less evident in the smaller size groups.

Figure 2A-2 shows the trend in the average number of chemical patents per firm in each sales group over the period 1960-1980 in the data set. As mentioned before, observations before 1968 and after 1977 are strongly biased downward because of the definition of the data set in terms of patents granted between 1968 and 1980.

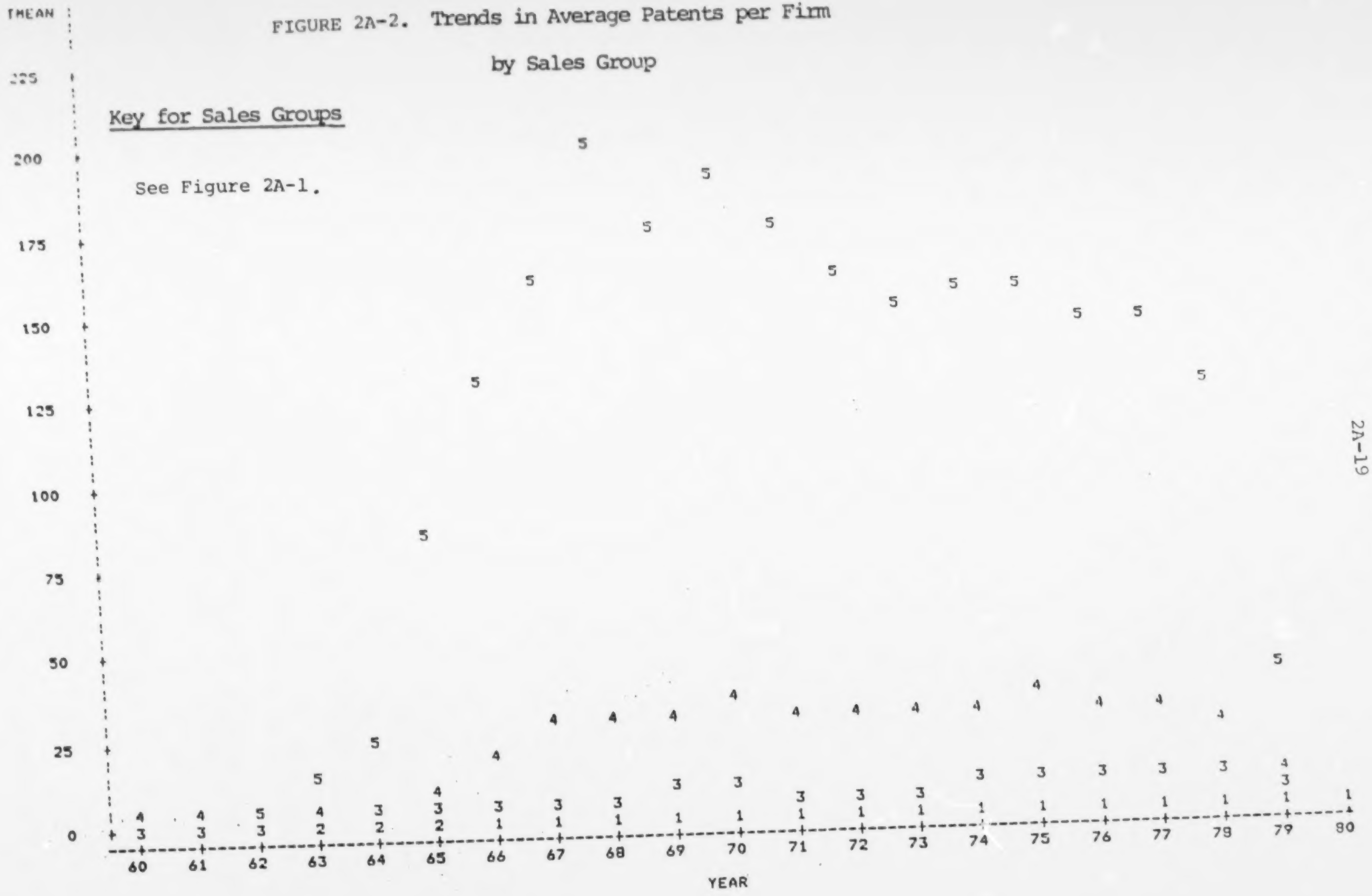
TABLE 2A-9.

1968-1978 TOTAL CHEMICAL PATENTS BY YEAR AND SALESGROUP

YEAR FREQUENCY PERCENT	SALESGRP .	UNDER 10	10-99	100-999	1000-2499	OVER 2500	TOTAL
68	5	0	10	160	556	1804	2530
	.	0.00	0.04	0.65	2.25	7.30	10.24
69	6	6	7	213	565	1572	2363
	.	0.02	0.03	0.86	2.29	6.36	9.56
70	8	0	6	215	629	1731	2581
	.	0.00	0.02	0.87	2.55	7.01	10.45
71	12	1	7	201	593	1592	2394
	.	0.00	0.03	0.81	2.40	6.44	9.69
72	10	2	7	179	538	1446	2172
	.	0.01	0.03	0.72	2.18	5.85	8.79
73	31	1	18	203	602	1352	2176
	.	0.00	0.07	0.82	2.44	5.47	8.81
74	13	0	9	253	585	1390	2237
	.	0.00	0.04	1.02	2.37	5.63	9.05
75	20	2	12	261	626	1376	2277
	.	0.01	0.05	1.06	2.53	5.57	9.22
76	13	1	14	238	534	1297	2084
	.	0.00	0.06	0.96	2.16	5.25	8.44
77	13	3	14	238	528	1308	2091
	.	0.01	0.06	0.96	2.14	5.29	8.46
78	15	1	5	211	445	1138	1800
	.	0.00	0.02	0.85	1.80	4.61	7.29
TOTAL	.	17	109	2372	6201	16006	24705
	.	0.07	0.44	9.60	25.10	64.79	100.00

FIGURE 2A-2. Trends in Average Patents per Firm

by Sales Group



Key for Sales Groups

See Figure 2A-1.

TABLE 2A-10.

1968-78 TOTAL CHEMICAL PATENTS BY MAJOR SIC CODE AND SALESGROUP

MAJSIC SALESGRP

FREQUENCY PERCENT	.	UNDER 10	10 - 99	100 - 999	1000-2499	OVER 2500	TOTAL
2810	17	0	24	477	5595	15071	21167
	.	0.00	0.10	1.93	22.65	61.00	85.68
2820	0	0	15	476	0	0	491
	.	0.00	0.06	1.93	0.00	0.00	1.99
2840	0	0	42	319	42	935	1338
	.	0.00	0.17	1.29	0.17	3.78	5.42
2850	0	0	10	140	109	0	259
	.	0.00	0.04	0.57	0.44	0.00	1.05
2860	0	0	0	269	0	0	269
	.	0.00	0.00	1.09	0.00	0.00	1.09
2870	0	0	0	0	73	0	73
	.	0.00	0.00	0.00	0.30	0.00	0.30
2890	129	17	18	629	96	0	760
	.	0.07	0.07	2.55	0.39	0.00	3.08
9999	0	0	0	62	286	0	348
	.	0.00	0.00	0.25	1.16	0.00	1.41
TOTAL	.	17	109	2372	6201	16006	24705
	.	0.07	0.44	9.60	25.10	64.79	100.00

TABLE 2A-11.
Mean and Standard Deviation of 1968-78 Chemical Patents per Million Dollars of R&D by Major SIC Code and Sales Group

	Sales Group					
Major SIC Code	Missing	Under 10	10-99	100-999	1000-2499	Over 2500
2810	1.293 (1.937)	---	0.319 (0.506)	1.414 (1.502)	2.890 (1.661)	2.251 (1.007)
2820	---	---	11.761 (10.181)	1.914 (0.469)	---	---
2830	---	---	---	---	---	---
2840	---	---	0.947 (1.922)	1.033 (1.188)	0.136 (0.198)	1.308 (0.934)
2850	---	---	1.808 (1.829)	0.993 (0.888)	0.682 (0.487)	---
2860	---	---	---	1.600 (0.911)	---	---
2870	---	---	---	---	0.944 (0.551)	---
2890	---	11.698 (26.467)	0.450 (0.944)	2.965 (3.920)	0.847 (0.242)	---
2910	---	---	---	---	---	---
9999	---	---	---	0.713 (0.728)	2.115 (0.821)	---

NA : Not applicable

--- Not available

( ) Standard deviation

Table 2A-10 shows the distribution of total chemical patents by major SIC code and sales group. Eighty-four percent of the total number of patents is generated by firms in SIC group 281 with sales greater than $1000 million. Another 8.5 percent of patents is contributed by SIC groups 284 and 289.

Table 2A-11 shows the mean and standard deviation of the ratio of chemical patents to R&D expenditures (patents per million dollars) for each major SIC code and sales group. There is a wide range of values, from .136 for SIC group 284 with $1000-2499 million in sales to 11.761 for SIC group 282 with $10-99 million in sales. Much of this variability results from the small number of firms in some cells, as was shown in Table 2A-5. It is difficult to make any generalizations from this distribution.

Figure 2A-3 shows the path of the average ratio of chemical patents to R&D expenditures over the period 1968-1798 for each sales group. The average values for the smallest size group (under $10 million) are much greater than those of the others, and also are much more erratic, reflecting the small sample size. The data suggest a slight downward trend for the largest two size groups over the period and a rise and then a decline for the $100-999 million sales group, while the trends for the smallest size groups are more difficult to detect.

Product/Process Ratio

Table 2A-12 shows the mean and standard deviation of the share of product patents in total chemical patents by major SIC code and sales group. There

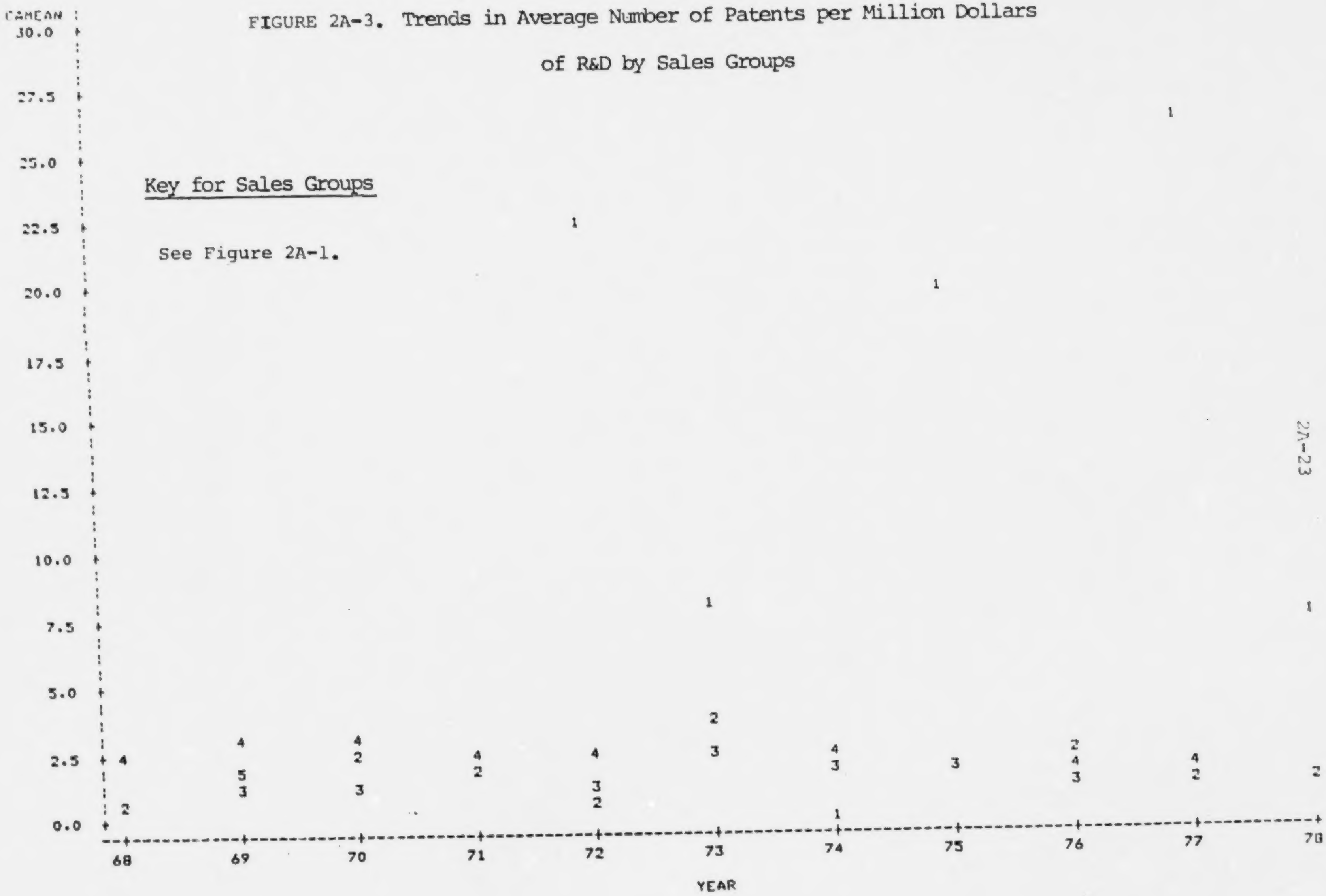


FIGURE 2A-3. Trends in Average Number of Patents per Million Dollars of R&D by Sales Groups

Key for Sales Groups

See Figure 2A-1.

TABLE 2A-12.

Mean & Standard Deviation of 1968-78 Chemical Product Patent Share by Major SIC Code and Sales Group

	Sales Group					
Major SIC Code	Missing	Under 10	10-99	100-999	1000-2499	Over 2500
2810	0.711 (0.393)	---	0.635 (0.427)	0.390 (0.375)	0.578 (0.160)	0.549 (0.137)
2820	---	---	0.380 (0.415)	0.725 (0.081)	---	
2830	---	---	---	---	---	---
2840	---	---	0.683 (0.404)	0.697 (0.333)	0.584 (0.419)	0.802 (0.095)
2850	---	---	0.813 (0.239)	0.606 (0.345)	0.638 (0.217)	---
2860	---	---	---	0.774 (0.250)	---	---
2870	---	---	---	---	0.448 (0.231)	---
2890	0.585 (0.291)	0.625 (0.489)	0.923 (0.188)	0.527 (0.342)	0.633 (0.175)	---
2910	---	---	---	---	---	---
9999	---	---	---	0.276 (0.369)	0.658 (0.084)	---

NA: Not applicable

--- Not available

( ) Standard Deviation

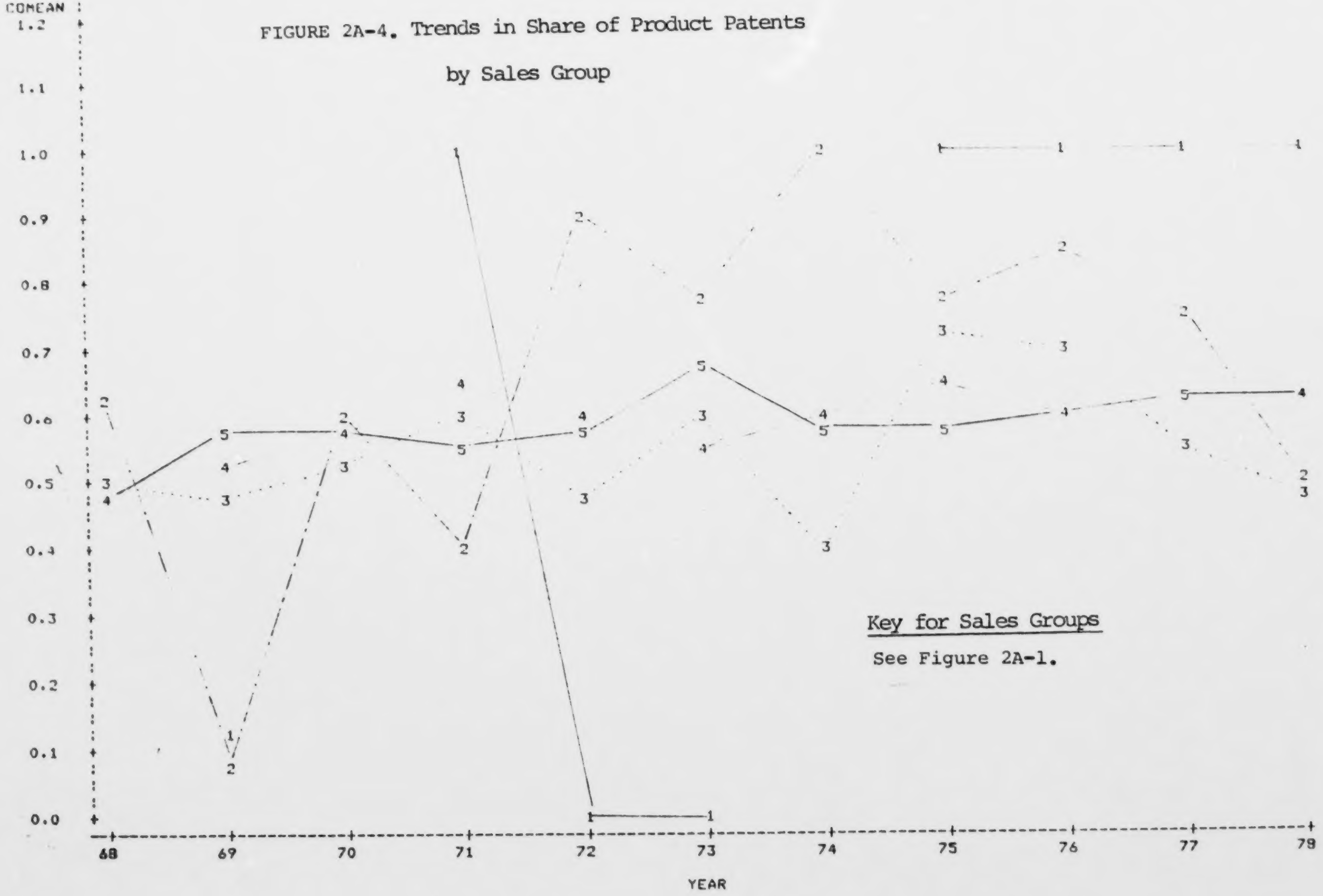


FIGURE 2A-4. Trends in Share of Product Patents

by Sales Group

Key for Sales Groups

See Figure 2A-1.

appear to be significant differences across SIC groups, but more explicit tests are needed to confirm this (see below).

Figure 2A-4 shows the trend in the average fraction of product patents per firm in each sales group over the period 1968-1978., For firms with sales greater than $100 million, the product ratio values are bounded between .4 and .7, with most values falling between .5 and .6. Overall there may be a slight increase in the product share over the period, and there is no consistent pattern of differences among the three largest sales groups. The results for the two smaller sales groups are much more erratic. The smallest sales group value is based on observations on only one firm, and the second sales group value is based on observations on five to seven firms.

Analysis of Variance

The contributions of across time and across company variance to explaining the total variance of the ratio of R&D to sales, the ratio of chemical patents to R&D expenditures, and the ratio of product to total patents were investigated using analysis of variance techniques.

Table 2A-13 shows the results for the R&D to sales ratio. Inter-firm differences explain 80 percent of total variance, and across time differences contribute an added 1.1 percent. A separate estimate with only time included explained 3.3 percent of the total, and had an F value (21, 1110) of 1.82, significant at the 2 percent level.

TABLE 2A-13.

```
                                   GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: RDDSALE

SOURCE                 DF       SUM OF SQUARES       MEAN SQUARE      F VALUE        PR > F      R-SQUARE          C.V.

MODEL                  21           0.03040840        0.00144802         1.82        0.0131      0.033330      102.4698

ERROR                1110           0.88193291        0.00079453                    STD DEV                RDDSALE MEAN

CORRECTED TOTAL      1131           0.91234131                                   0.02818748                  0.02750808

SOURCE                 DF            TYPE I SS     F VALUE     PR > F        DF         TYPE IV SS     F VALUE     PR > F

YEAR                   21           0.03040840        1.82     0.0131        21         0.03040840        1.82     0.0131
```

RDDSALE = R&D / Sales

TABLE 2A-14.

```
                                   GENERAL LINEAR MODELS PROCEDURE
DEPENDENT VARIABLE: CPATDRD

SOURCE                 DF       SUM OF SQUARES        MEAN SQUARE       F VALUE         PR > F       R-SQUARE          C.V.

MODEL                  86       8922.37622554        103.74856076          4.27         0.0001       0.396220      219.8863

ERROR                 559      13596.34394878         24.32261887                      STD DEV                 CPATDRD MEAN

CORRECTED TOTAL       645      22518.72017432                                        4.93179672                   2.24288540

SOURCE                 DF          TYPE I SS      F VALUE       PR > F        DF          TYPE IV SS       F VALUE       PR > F

COMPANY                76      8801.05163778         4.76       0.0001
YEAR                   10       121.32458776         0.50       0.8911        10        121.32458776          0.50       0.8911
```

CPATDRD = Chemical Patents / R&D

TABLE 2A-15.

```
                                GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: CPRODFR

SOURCE                 DF     SUM OF SQUARES       MEAN SQUARE      F VALUE         PR > F      R-SQUARE          C.V.

MODEL                  88        26.34752200        0.29940366         5.28         0.0001      0.467795       40.5606

ERROR                 529        29.97524685        0.05666398                     STD DEV                 CPRODFR MEAN

CORRECTED TOTAL       617        56.32276885                                    0.23804198                   0.58687930

SOURCE                 DF          TYPE I SS     F VALUE     PR > F       DF           TYPE IV SS     F VALUE     PR > F

COMPANY                78        24.99193667        5.65     0.0001
YEAR                   10         1.35558533        2.39     0.0088       10           1.35558533        2.39     0.0088
```

CPRODFR = Chemical Product Patents / Total Chemical Patents

Table 2A-14 shows the results for the ratio of chemical patents to R&D expenditures. This analysis was done for the period 1968-1978 because of the problem of incomplete reporting of patent applications outside this period. Inter-firm variance explains 39.1 percent of the total and across time variance adds another 0.5 percent. This is a significant reduction from the amount of explanatory power for the R&D to sales ratio.

Table 2A-15 shows the results for the share of product patents in total chemical patents for the period 1968-1978. Inter-firm variation explains 44.3 percent of the total and time contributes another 2.4 percent.

To sum up, inter-firm variance is an important factor in explaining all three variables, but there is considerably more residual variance in the patent relationships. The time variable is only marginally significant in explaining overall variance.

The effect of the major SIC codes for each firm on the ratio of R&D to sales, the number of chemical patents per million dollars of R&D, and the share of product patents was also tested. The results are shown in Table 2A-16. In all three cases, the F values of the models were significant, but the percentage of explained variance was relatively small -- 11.6 percent for the R&D to sales ratio, 5.5 percent for the ratio of chemical patents to R&D expenditures, and 7.5 percent for the share of product patents.

TABLE 2A-16.

```
                     GENERAL LINEAR MODELS PROCEDURE
DEPENDENT VARIABLE: RODSALE

SOURCE             DF     SUM OF SQUARES      MEAN SQUARE      F VALUE       PR > F      R-SQUARE        C.V.

MODEL               9         0.06152848       0.00683650        11.85       0.0001      0.116944     91.9329

ERROR             805         0.46460728       0.00057715                   STD DEV               RODSALE MEAN

CORRECTED TOTAL   814         0.52613576                                 0.02402399                 0.02613209

SOURCE             DF          TYPE I SS    F VALUE     PR > F      DF        TYPE IV SS    F VALUE     PR > F

MAJSIC              9         0.06152848      11.85     0.0001       9        0.06152848      11.85     0.0001



DEPENDENT VARIABLE: CPATORD

SOURCE             DF     SUM OF SQUARES      MEAN SQUARE      F VALUE       PR > F      R-SQUARE        C.V.

MODEL               7      1243.27232436     177.61033205         5.33       0.0001      0.055210    257.4761

ERROR             638     21275.87742761      33.34777026                   STD DEV               CPATORD MEAN

CORRECTED TOTAL   645     22519.14975197                                 5.77475283                 2.24283058

SOURCE             DF          TYPE I SS    F VALUE     PR > F      DF        TYPE IV SS    F VALUE     PR > F

MAJSIC              7      1243.27232436       5.33     0.0001       7     1243.27232436       5.33     0.0001



DEPENDENT VARIABLE: CPRODFR

SOURCE             DF     SUM OF SQUARES      MEAN SQUARE      F VALUE       PR > F      R-SQUARE        C.V.

MODEL               7         4.20801742       0.60114535         7.04       0.0001      0.074713     49.8043

ERROR             610        52.11475143       0.08543402                   STD DEV               CPRODFR MEAN

CORRECTED TOTAL   617        56.32276885                                 0.29229098                 0.58687930

SOURCE             DF          TYPE I SS    F VALUE     PR > F      DF        TYPE IV SS    F VALUE     PR > F

MAJSIC              7         4.20801742       7.04     0.0001       7        4.20801742       7.04     0.0001
```

Appendix 2-B.
Test of Reliability of Patent Titles as a Means to Classify Patents.

In order to develop a data base which identifies product and process patents, patent titles provided by the U.S. Patent Office have been used to classify patents. Because of questions about the reliability of patent titles for this purpose, an experiment was conducted to compare the patent classification obtained using the patent title alone with the one resulting from the more complete information provided by the patent abstract, for a sample of 108 patents. This memorandum describes that experiment and its results.

Using the general classes of product vs. process and chemical-related vs. unrelated, use of the patent title alone produced mis-classifications in 13 percent of the cases. Using a more specific set of classifications based on distinguishing between chemical compounds and chemical compositions and their related processes resulted in 30 percent mis-classifications. We believe that the 13 percent is more relevant, since we are concerned mainly with the broad split between products and processes.

Classification Scheme

Initially, the patents were grouped into six categories:

1. chemical products;
2. chemical processes;
3. chemical products and processes;
4. non-chemical patents;
5. new uses; and
6. unknown.

Based on subsequent discussion with EPA, it was decided to focus on the product/process distinction rather than on the chemical/non-chemical distinction, since TSCA could affect both the latter types. In order to allow both kinds of comparisons to be made, the following classification scheme was used:

1. Chemical Substances; e.g., resins, esters;
2. Chemical Processes; e.g., reaction, recovery;
3. Both Chemical Substance and Processes; e.g., chemical product and process for making same;

4a. Products that are not chemical substances; e.g., composition;

4b. Processes that are not chemical processes; e.g., coating, laminating;

4c. Products and Processes that are non-chemical;

4d. Products and processes not directly related to any chemical product or process;

5. New uses of chemical substances; and
6. Unknown.

This was the classification scheme used in the experiment.

This classification can be transformed into a much simpler classification scheme of product, process, non-chemical, and unknown by the following mapping:

General Class	Specific Class
1. Products	1, 3, 4a, 4c, 5
2. Process	2, 4b
3. Non-chemical	4d
4. Unknown	6

It is this general classification that is used in the analysis of Chapters 3 and 4, with the first two groups used to proxy for product and process innovation, and the latter two groups ignored. Results of the experiment are presented for kinds of classifications, although we believe the results for the general class to be more relevant for the study.

Methodology

For a sample of patents, patent abstacts and claim descriptions were obtained from the claims/U.S. Patents data base available through the Dialog data retrieval service. These patents were classified using the specific classes described above. (Pains were taken to ensure that the classifiers did not see the original classifications by title during the course of the experiment.) The classifications were then compared with the classifications by title and the results tallied.

Selection of Sample

A sample totalling 108 patents was obtained from one representative large chemical firm (Allied Chemical, 50 patents) and two medium-size companies (Crompton and Knowles, 27 patents, and Dow Corning, 31 patents). The companies were chosen because they had a broad mix of patent types. The individual patents were chosen from uniform intervals within each company listing to provide a sample with the ratio of number of patents from the three companies of 2:1:1, respectively.

Results

Using the specific patent classes, 32 out of 108 (30 percent) received different classifications using patent title alone rather than the abstract. This result becomes less alarming when the general patent classes are used; the number of mismatches drops to 14 out of 108 (13 percent). Of the 14, eight were incorrectly classified as product or process, and six were incorrectly classified as chemical or non-chemical. In addition, three of the 14 mismatches occurred because, even on the basis of the title alone, the patents were mis-classified on the first round.

Chapter 3

Knowledge Production Function

Introduction

As described in the preceding chapter, the innovation model developed in this report consists of three relationships: a knowledge production function, a revenue function, and an R&D demand function. The knowledge production function relates R&D inputs to the creation of "an increment of knowledge" such as a new chemical entity (NCE) or a new process. Increments to knowledge are not directly observable; however, patents are a useful proxy for innovation output. A given increment of knowledge typically results from R&D expenditures over a number of years. This process may be modeled using the concept of an R&D "stock" which summarizes the effect of expenditures over several years. In turn, increments to knowledge by the firm over a number of years may be considered a "stock of technical knowledge" which affects its overall productivity. Following a general description of the model, three sets of results are presented.

General Model

The R&D process may be conceived of as the use of various scientific and technical inputs to develop a body of technical knowledge which the firm uses to maintain and enhance its production possibilities.* For the chemical industry, typical additions to this body of knowledge (innovation) consist of new chemical entities (NCE's), new uses for existing chemicals, new processes, and cost savings for existing processes. (The first two kinds of innovation are subject to TSCA regulations while the last two are not.) The knowledge production function attempts to relates measures of R&D inputs to a measure of additions to knowledge. The methodology used here follows that developed by Griliches, Schankerman and Pakes, among others. Rather than focusing on individual projects, it attempts to model the average relationship between a firm's R&D inputs and its innovation outputs.

---

*In practice, the line between R&D and commercialization is fuzzy, since the innovation does not begin to benefit the firm until the commercialization process is complete.

Following Pakes and Griliches*, K is defined as the level or stock of economically valuable technical knowledge, where $\dot{K}$ is the net accretion to K per unit of time. The model assumes that K is produced by current and past research and development activities:

$$K = f(\overline{R}) + e \qquad (3\text{-}1)$$

where $\overline{R}$ is the "stock" of R&D and $f(\overline{R})$ is the "technical knowledge production function." The error term (e) reflects the effects of other variables neglected in the analysis. If R&D is assumed to decay at a constant exponential rate, the stock of R&D can be measured by a combination of R&D expenditures in different years, i.e., a lag structure of R&D expenditures such as

$$\overline{R} = R_0 + (1-d)R_{-1} + (1-d)^2 R_{-2} + (1-d)^3 R_{-3} \qquad (3\text{-}2)$$

where:

R_i = R&D expenditures in year i, and

d = the R&D depreciation or decay rate.

---

* This section is in large part based on "Patents and R&D at the Firm Level: A First Look" by Ariel Pakes and Zvi Griliches, National Bureau of Economic Research, Inc., Working Paper No. 651, October 1981.

The use of several years reflects the fact that a particular innovation results from R&D expenditures over a number of years. The notion that a firm's technical knowledge decays over time is discussed by Griliches (1979). For example, if only current R&D affects the production of technical knowledge, then

$$\bar{R} = R_0$$

The stock of knowledge, K and its net accretion, $\dot{K}$, are unobservable variables; they must be measured with surrogate indicators. Several candidate indicators of $\dot{K}$ including the number of new chemical entities, new chemical sales intensity, the number of new patents, weighted patent counts, and the number of significant innovations were considered. This report uses patent data because they are more readily obtainable than the other innovation indicators.

There are problems in using patent data as the indicator of $\dot{K}$. The first problem relates to the propensity of a firm to patent its inventions. That propensity (as measured, say, by the number of patents per patentable invention) may change over time as the firm changes its marketing, competitive, or R&D strategy. If the propensity to patent changes significantly, patents will be a poor measure of $\dot{K}$. Differences in the propensity to patent among firms also make patent data less reliable.

The second problem in using patent data relates to differences among inventions. Some patents are for important breakthroughs in technology and others are for marginal changes in products or processes. Both variations in the propensity to patent and variations in the importance of individual patents mean that patents are an error-ridden measure of increments to knowledge.

Since $\dot{K}$ is an unobservable variable, the relationship must be estimated using an observable variable, P, which is a surrogate for $\dot{K}$. This can be expressed as:

$$P = g(\dot{K}) + u \tag{3-3}$$

where u is an error term and g( ) the functional relationship between P and $\dot{K}$. Thus the model has the two equations:

$$\dot{K} = f(\overline{R}) + e$$

$$P = g(\dot{K}) + u$$

The specific functional form of both the knowledge production function and the indicator function can be hypothesized, but since $\dot{K}$ is unobservable the following function is estimated:

$$P = h(\overline{R}) + n \tag{3-4}$$

This function thus combines the effects of the knowledge production function and the indicator function, and the error term n contains the effects of both error terms, e and u.

The functional forms used in this patent equation research are

$$P = a\overline{R}^{\beta} \tag{3-5}$$

and

$$P = a_t \overline{R}^{\beta} \tag{3-6}$$

where:

P = the number of patents

$\overline{R}$ = the stock of R&D

In the latter form, the productivity of R&D, represented by the efficiency factor a, is allowed to vary over time.

In all of the functions estimated, the method of determining the lag structure of the R&D stock must be chosen. Two approaches are possible. The first is to assume a particular form for the lag structure and then estimate the equations using that form. The second is to use the estimating procedure to derive a lag structure.

Estimates Based on Aggregate Patent Data

The basic model of the relationship between firm R&D expenditures and total patent applications is:

$$\log P_t = a + b \log R_t + e_t \tag{3-7}$$

where:

a = constant term
P_t = patent applications in year t
R_t = some measure of R&D expenditures in year t, and previous years
e_t = error term.

The basic data set consists of 436 observations for 63 firms. The analysis addresses the following questions: the nature of the lag relationship between patents and R&D applications; the relative contributions of across firm and within firm variations in explaining patenting behavior; and the effect of different industry groups, as defined by SIC codes, on the elasticity of patents with respect to R&D.*

All the equations are estimated using ordinary least squares (OLS) techniques. Most of the estimation work was done on a sample of 63 firms which have positive values for both patent applications and R&D expenditures. The data include patent applications for which patents were granted by the U.S. Patent Office between 1969 and 1980. This definition causes observations before about 1968 and after 1978 to be incomplete, so the analysis was done only for data falling within this time period.

Alternative Specifications of R&D Effect

Several different specifications of the effect of R&D expenditures on patenting were tried and compared: current year R&D alone, a stock of R&D constructed using the average growth rate of R&D expenditures and a fixed depreciation rate, and an unrestricted lag structure for the most recent five years of R&D. As is usual in patent-R&D equations, the various specifications differed little in explanatory power.

Table 3-1 shows the effect of current year R&D with and without time dummies and of an R&D stock with zero lags. The F-test shows the time dummies to be not significant. The coefficient for current-year R&D alone is .900 (.026). This is high compared with other values reported by Pakes and Griliches (1980), .61 (.08); and Hausman, Hall, and Griliches (1984), .81 (.02). The variability of the results seems to be due to two main sources:

---

* Additional results which address the effect of zero or missing observations of patents and R&D expenditures and the effect of the size of R&D expenditures are presented in Appendix 3-A.

Table 3-1.

Log Patent Regression Results - Flow vs. Stock Effects
Observations = 436

Variables	1	2	3
Log R&D	0.900 (.026)	0.886 (.027)	
Log R&D Stock			0.860 (.029)
Other Variables Included	Intercept	Intercept, Year Dummies	Intercept
Standard Error	0.805	0.801	0.765
R^2	0.7304	0.7378	0.7264
F-test		$F_{11,424} = 1.10$ Year dummies	

Figures in parentheses are estimated standard errors.

the proportion of zero values of patents in the sample, and the size range of the sample. (The former effect is examined in Appendix 3-A.) It may also be affected by the fact that only chemical firms are included, while the other studies included all kinds of manufacturing firms.

Equation (3) of Table 3-1 shows the elasticity for the constructed R&D stock. Its value of .860 (.028) is not significantly different from that of the current year R&D, and the R^2 is slightly lower. This evidence can be interpreted as supporting the view that the patent application step takes place very early in the R&D process, so that lagged effects are not significant. On the other hand, the difference is very small, so the results may simply be due to lack of variability in the R&D data and patent data.

Table 3-2 shows the effect of lagging the R&D stock by 0, 1, and 2 years.* The results show a consistent though small decline in explanatory power, again suggesting that the contemporaneous specification is preferred. The elasticity estimate also declines slightly, which is consistent with increased noise in the relationship, given the longer time lag.

Table 3-3 again examines the question of lagged effects by comparing an unrestricted five year lag structure on R&D expenditures with current year R&D alone and with the R&D stock.* A comparison of R^2's shows that the unrestricted lag structure has only slightly greater explanatory power. The F-test shows that the lag structure is significant as a whole, but not to any great degree. Also, the sum of the coefficients for the unrestricted lag form is .975, almost identical to the value for current year R&D alone. However, the lag structure implied by these coefficients, with so much of the weight on the earliest R&D, does not make a lot of sense. This may be the

---

* The sample size is smaller than in Table 3-1 because of the lagged values.

Table 3-2.

Log Patent Regression Results - Lag Effects
Observations = 361

Variables	1	2	3
Log R&D Stock	0.884 (0.029)		
Log R&D Stock_{-1}		0.870 (.029)	
Log R&D Stock_{-2}			.854 (.029)
Other Variables Included	Year Dummies	Year Dummies	Year Dummies
Standard Error	0.765	0.776	0.794
R^2	0.7880	0.7818	0.7801

Figures in parentheses are estimated standard errors.

Table 3-3.

Log Patent Regression Results - Lag Effects (continued)

Observations = 303

Variables	1	2	3
Log R&D	.974 (.035)	-0.062 (.267)	
Log $R\&D_{-1}$		0.838 (.396)	
Log $R\&D_{-2}$		-0.473 (.420)	
Log $R\&D_{-3}$		-0.078 (.422)	
Log $R\&D_{-4}$		0.750 (.306)	
Log R&D Stock			.969 (.033)
Other Variables	Intercept	Intercept	Intercept
Standard Error	.855	0.792	.795
R^2	.7188	0.7429	.7384
F-test		$F_{4,297} = 6.94$ Lagged R&D	

Figures in parentheses are estimated standard errors.

result of fixed effects problems or of truncation problems, i.e. a longer time lag should have been used.

These results indicate that there is little support for strongly preferring one specification of R&D effects over another, and that little is lost by using current R&D or a current R&D stock alone.

Within Firm Variation

Table 3-4 shows similar specifications to those before, except the dependent and independent variables are defined as deviations from company means. This has the effect of eliminating all firm-specific behavior from the data. The regressions are done with current R&D alone or the R&D stock and with and without time dummies. For current R&D, the elasticity estimate is considerably lower than in the overall regressions (Table 3-1), falling from about .90 to .35-.40. The stock elasticity is affected less; in the last column, with time dummies included, it is essentially unchanged from the previous results. Also, the stock form (with time dummies) has considerably higher explanatory power than the equivalent form using the R&D flow. Finally, the time dummies are more significant here than they were in Table 3-1. This is because they cause about the same absolute reduction in the residual sum of squares, leading to a larger relative reduction (and hence a larger F-statistic) since the within residual sum of squares is much smaller than the overall sum of squares.

These results show that the between-firm variation in the average propensity to patent explains patenting behavior much more than within firm variation, and estimates of the elasticity of patenting with respect to R&D are generally lower in the within dimension. Unfortunately, it is difficult to know whether it is the overall or within results that are closer to being "right." This is consistent with other studies.

Table 3-4.

Log Patent Regression Results - Deviations from Firm Means
Observations = 436

Variables	1	2	3	4
Log of R&D	.355 (.085)	.394 (.088)		
Log of R&D Stock			0.514 (.121)	0.961 (0.147)
Other Variables Included		Year Dummies		Year Dummies
Standard Error	0.153	0.146	0.152	0.139
R^2	0.0383	0.1015	0.0397	0.1453
F-test		$F_{10,425}$ = 3.00 year dummies		$F_{10,425}$ = 5.25 year dummies

Figures in parentheses are estimated standard errors.

Industry Groups

Two sets of data on industry groups were available for firms in the sample. The Disclosure data base assigns each firm to a unique major 3-digit SIC code. In addition, Dun & Bradstreet's Million Dollar Directory identifies all major 4-digit SIC activities at each firm. Separate equations were estimated for each type of industry classification. Separate dummy variables were estimated for each 3- or 4-digit SIC code falling within the scope of the study, while other SIC's were lumped together. The results are shown in Table 3-5.

Both kinds of industry group variables have F-statistics which clearly reject the null hypothesis. The 4-digit SIC's (equation (3)) have greater explanatory power than the 3-digit SIC's (equation (2)), but there are more of them -- 31 separate groups versus eight. The results also show that as the specificity of the industry groups increases, the elasticity estimates decline in absoluate value and in significance. This is consistent with the results for the within firm estimates, since that specification corresponds to creating a separate dummy variable (i.e., group) for each firm. In other words, increasing the specificity of the industry groups tends to reduce the ability of the residual variability of R&D expenditures within the group to explain the variability of patenting behavior within that group.

The interpretation of this pattern of results is clouded by the failure of industry effects to be important in the line of business results discussed in the next section. As will be discussed in Chapter 6, one item for future research would be to utilize the various levels of industry aggregation that are available, to model in more detail the structure of industry and firm effects.

Table 3-5.

Log Patent Regression Results - Industry Groups
Observations = 436

Variables	1	2	3	4
Log R&D	0.886 (0.027)	0.785 (0.028)	0.698 (0.032)	
Other Variables Included	Year Dummies Intercept	Year, 3-digit SIC Dummies	Year, 4-digit SIC Dummies	Year, 4-digit SIC Dummies, 4-digit SIC slope terms
Standard Error	0.801	0.581	0.424	.324
R^2	0.7378	0.8130	0.8708	.9086
F-test		$F_{7,417} = 23.9$	$F_{30,394} = 18.3$	$F_{28,366} = 5.41$
		3-digit SIC Codes	4-digit SIC Codes	4-digit SIC slope terms

Figures in parentheses are estimated standard errors.

Equation (4) shows the effect of adding separate slope terms for R&D for each 4-digit SIC industry to equation (3). The overall F-value is 5.41 (28,366), which is significant at the usual confidence levels. Only six of the 28 individual slope terms are significant at the five percent level.

Table 3-6 shows qualitatively the effects of the major industry groups on average patenting behavior. The results for the 4-digit SIC groups were aggregated to the 3-digit level based on the average number of coefficients in each group which were positive or negative and significant or not significant. Fortunately, the results of the two data sets are largely consistent with one another. SIC groups 281, 282, 283, and 291 have positive signs while 284, 285, 287 and 289 have negative signs. The sign for 286 varies between the two data sets but is not significant at the 4-digit level. This result is important because it shows that firm level data can be used to make inferences about activity at a more disaggregated level within the firm. It allows us to make fuller use of the data, because many more firms report activities in each 4-digit SIC group than have their major SIC in that group.

Estimates Based on the FTC Line of Business Data*

This section describes the results of the econometric analysis of the relationship between patenting activity and R&D expenditures at the line of business (LB) level. The R&D data comes from the Federal Trade Commission's (FTC) Line of Business Survey, with adjustments made by F.M. Scherer. The patent data were matched to the LB's by F.M. Scherer for another research

---

* Certification: This section has been reviewed by the Manager of the Federal Trade Commission's Line of Business Program for identification of individual company LB data, its release will not result in such disclosure.

Table 3-6.

Effect of SIC Group on Average Patenting Behavior

SIC Code	3-digit Major SIC*	4-digit SIC**
281	+ (0)	+ (+)
282	+ (0)	+ (+)
283	***	+ (0)
284	- (+)	- (+)
285	- (+)	- (+)
286	- (+)	+ (0)
287	- (+)	- (+)
289	- (+)	- (0)
291	***	+ (+)

+ Indicates positive coefficient.

- Indicates negative coefficient.

(0) Indicates not significant at 5 percent level.

(+) Indicates significant at 5 percent level.

* Table 3-5, Equation 2.

** Table 3-5, Equation 3.

*** No observation in this cell with both patent and R&D data.

project.* Only 1974 data are included in the patent - R&D file used in this study.**

The LB data are valuable because they are disaggregated to (roughly) the 4-digit SIC level, and hence pinpoint much more precisely the chemical activities of firms which are the main focus of this study. In contrast, firm-level data are contaminated by the broad range of activities that many large firms engage in, which makes it difficult to draw conclusions about chemical industry behavior, both overall and within particular segments. The major limitations of the LB data are that it includes mainly very large firms (sales greater than $100 million) and covers only the years 1974-1976.

Because the matched patent and R&D data are available for only one year, it was not possible to examine questions of the lag structure between R&D and patents. However, previous work indicates that even fairly long time series data do not shed much light on these questions, so the loss is not great. More importantly, however, the lack of a time series makes it impossible to estimate the relative contributions of firm-specific and time-varying behavior in explaining patenting activity.

Industry-Specific Behavior

Table 3-7 compares the results of a simple log-log equation which relates patents to R&D (1), to the results which include industry dummies for individual LB codes (2). Equation (1) yields a slope coefficient (elasticity)

---

* See Scherer (1984).

** A more complete description is presented in Chapter 2.

Table 3-7.

Log Patent Regression: Industry Specific Behavior

Observation = 257

Variables	1	2
Log R&D	.655 (.044)	.616 (.046)
Other Variables Included	Intercept	Intercept, industry dummies
Standard Error	.907	.789
R^2	.471	.565
F-test	--	$F_{14,241} = 3.71$ Industry dummies

Figures in parentheses are estimated standard errors.

of .655, which is significantly lower than that obtained in the firm-level results (.85 to .90). Also the R^2 is lower, .471 versus .70 or greater. Both patents and R&D are likely to be measured with more error at the LB level than at the firm level, thus resulting in both a poorer fit and a smaller coefficient on R&D. It must be noted that all observations with zero R&D or zero number of patents are dropped from this and next four regressions because of the log-log equation.

Surprisingly, the industry effects are not very strong. Including industry dummies in equation (2) yields an F value which is only marginally significant, and reduces the elasticity of patents by less than 10 percent. In contrast, 4-digit SIC dummies were highly significant in the firm-level regressions, and had a much greater effect on the slope coefficient (reducing it from .886 to .680). This suggests that the strong results for the firm-level data may be spurious. For example, the industry dummies may be acting as proxies for firm-specific behavior, or for differences in average firm size. In this connection, it should be remembered that the LB data base contains mainly very large firms.

Comparison of Firm Level and Line of Business Results

Two additional issues, scale effects and the inclusion of zero or missing patent and R&D observations, were investigated. The results are presented in Appendix 3-A and summarized here. Because some companies or LB's report no patents in a given year while reporting R&D expenditures, and vice versa, the estimates may be biased if these zero observations are not included. Therefore, both the firm-level and the LB results were reestimated in the manner suggested in Bound et al (1982, p.33). This greatly increases the size of both samples. In both samples, the R&D coefficient drops substantially. For the LB data, observations with zero patents have an expected value of 1/3 patent.

Scale effects were investigated both in terms of a dummy variable to divide the sample and in terms of a quadratic form. In the firm level data, large firms have a higher elasticity of patenting under both specifications. When SIC dummies are included in the quadratic form, they are significant as a whole while the magnitude of the quadratic coefficient is reduced by 2/3. This indicates a positive relationship between scale and industry group. For the LB data, no scale effects were evident when a size dummy was used. Adding industry dummies did not change this result. However, the quadratic form with industry group dummies does display positive scale effects. The magnitude of the scale coefficients, when industry dummies are included, are quite close for the firm-level and the LB data.

Overall, the results for the LB analysis are similar to those for the firm-level analysis, but there are some tantalizing differences. The overall elasticity of patenting is lower for the LB data, but this probably reflects errors in measurement of R&D at the LB level. Adding industry-specific intercepts and a quadratic term makes the coefficient estimates between the two data bases much closer. Both analyses show a positive but small and not very significant relationship between the patenting elasticity and the scale of R&D expenditures and the sales size of the patenting unit.

The major riddle is the difference in importance of industry-specific effects. The industry assignment of the firm is important in explaining variations in patenting across firms, but at the LB level it seems to make little difference. Since the LB data, at least in principle, gives a better picture of the industrial position of the firm, this is difficult to reconcile with the usual notions of industry effects. One explanation is that the LB data are so bad that tests using them have no power. The overall quality of the LB results in this and the next two chapters lend some support for this view. An alternative explanation is that the apparent industry effects at the firm level are really firm-specific effects. This is testable in the LB data, and should be a focus for future work.

Estimates that Distinguish Between Product and Process Innovation

The pre-market notification and testing requirements of the Toxic Substances Control Act (TSCA) may differentially affect the cost of product and process innovation. Therefore, a version of the knowledge production function was formulated that distinguishes between product and process innovations.

The Model

First, a relationship between the firm's R&D input, on the one hand, and its output of product and process innovations on the other is specified. In its most general form, this relationship can be represented by a transformation function:

$$T(\bar{R}_{it}, \dot{K}_{1it}, \dot{K}_{2it}) = 0 \qquad (3\text{-}8)$$

where $\bar{R}_{it}$ is the vector of past and present R&D expenditures of firm i at the time t, $\dot{K}_{1it}$ is the increase in the process knowledge stock, and $\dot{K}_{2it}$ is the increase in the product knowledge stock. In practice, process and product patents are used as proxies for $\dot{K}_1$ and $\dot{K}_2$. To simplify, it is assumed that $T(\cdot)$ is separable in inputs and outputs so it can be written:

$$g(\bar{R}_{it}) = f(\dot{K}_{1it}, \dot{K}_{2it}) \qquad (3\text{-}9)$$

This is not necessarily an innocuous assumption, but it is standard in the multiple output production function literature. (See Diewert (1974) and Hasankamp (1976).)

In principle, the function g(R) involves some sort of distributed lag. As discussed earlier in this chapter, using current R&D rather than the complete lag structure is a reasonable approximation. That approach is adopted herein, i.e. $g(\bar{R}_{it}) = R_{it}$. This should not be taken to imply that there are no lags in the true process; it is simply that the data do not permit their identification.

The only remaining issue is the functional form for f(˙). The linear-quadratic form suggested by Diewert (1974) is used:

$$
\begin{aligned}
R_{it} = A &+ B_1 \dot{K}_{1it}^{1/2} + B_2 \dot{K}_{2it}^{1/2} + B_{11} \dot{K}_{1it} \\
&+ B_{22} \dot{K}_{2it} + B_{12} \dot{K}_{1it}^{1/2} \dot{K}_{2it}^{1/2}
\end{aligned}
\qquad (3\text{-}10)
$$

Equation (3-10) is the form of the KPF whose estimation is discussed below. Its advantages are that it is a local second order approximation to any well-behaved function, and it is defined for $\dot{K}_1, \dot{K}_2 = 0$, which is important for our data.

The derivation of the knowledge production function was done in a completely deterministic framework. In order to estimate the KPF, a stochastic structure must be specified. The easiest thing to do is to simply append the usual error term to Equation (3-10). This specification, however, has implications that must be considered. It means that patents are viewed as exogenous and R&D as stochastically determined. This seems counter-intuitive, but there does not seem to be an easy way around it. What it says is that the firm picks the output combination it desires and the KPF tells it the expected R&D necessary to produce those outputs. It then takes a random draw, which determines the actual R&D (sometimes more, sometimes less) that it spends. There is no error in the composition of output. In Figure 3-1, the firm aims for point A; the error term causes it to move in or out along the ray to 0A.

It might seem more plausible that in addition to this "efficiency error" there is an "output mix error" that would move the firm along the tangent to the frontier at point A. Some time was spent in trying to specify a model incorporating both errors, but it was too complicated for the present purposes. Therefore, Equation (3-10) was estimated simply by OLS.

The Results

The KPF was estimated using 638 observations on 77 companies. The distribution of observations by year is given in Table 3-8. Because Equation (3-10) is an approximation, the coefficients themselves are not directly interpretable as parameters of economic interest. The derivatives of the function f depend also on the values of $\dot{K}_1$ and $\dot{K}_2$ at which they are evaluated. These derivatives were evaluated as follows:

$$f_1 = B_1\dot{K}_1^{-1/2} + B_{11} + B_{12}\,(\dot{K}_2/\dot{K}_1)^{1/2}$$

$$f_2 = B_2\dot{K}_2^{-1/2} + B_{22} + B_{12}\,(\dot{K}_1/\dot{K}_2)^{1/2}$$

$$f_{11} = -.5\,(B_1\dot{K}_1^{-3/2} + B_{12}\dot{K}_2^{1/2}\,\dot{K}_1^{-3/2})$$

FIGURE 3-1. Knowledge Production Function

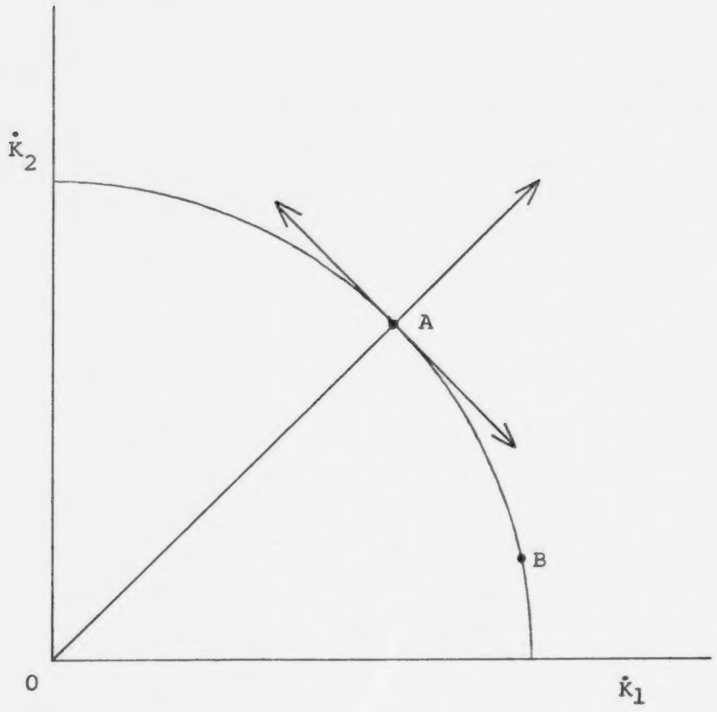

$\dot{K}_2$
A
B
0
$\dot{K}_1$

Table 3-8.

Distribution of Observations by Year For the Knowledge Production Function (KPF)

Year	Number of Firms
1968	26
1969	31
1970	38
1971	52
1972	55
1973	59
1974	72
1975	77
1976	77
1977	76
1978	75
Total	638

These include all observations in 1968 - 1978 for which R&D is greater than 0.

$$f_{22} = -.5 \quad (B_2\dot{K}_2^{-3/2} + B_{12}\dot{K}_1^{1/2}\dot{K}_2^{-3/2})$$

$$f_{12} = -.5 \quad (B_{12}\dot{K}_1^{-1/2}\dot{K}_2^{-1/2})$$

For the purpose of Table 3-9, all functions of $\dot{K}_1$ and $\dot{K}_2$ were evaluated at the mean of the function for the data. (Note that this required ignoring zeros for all functions involving reciprocals.) The column labelled "T" in Table 3-9 gives the parameter which determines whether the function f is quasi-convex. It is related to, and in fact determines the sign of, the elasticity of substitution. According to Allen (1938):

$$T = -(f_{11}f_2^2 - 2f_{12}f_1f_2 + f_{22}f_1^2), \text{ and}$$

$$s = \frac{(f_1f_2)(\dot{K}_1f_1 + \dot{K}_2f_2)}{(\dot{K}_1\dot{K}_2)T}$$

For inputs, the level sets of f are expected to be concave to the origin, which is the case if T (and hence s) is positive. For outputs, the level sets are expected to be convex; s and T are both negative. The border between convex and concave level sets is where T = 0 and s approaches either negative or positive infinity, depending on the direction of approach. Thus, the test of whether the KPF has the "right" curvature is the test of whether T is negative.

Table 3-9 shows, essentially, that the "total" results are fair, and the "within" results are poor. The total estimates imply, at the mean of the data, that the marginal rate of transformation (MRT) of K_2 to K_1 is about

Table 3-9.

Generalized Linear Quadratic Knowledge Production Function (KPF)

$$R = f(\dot{K}_1, \dot{K}_2) = A + B_1\dot{K}_1^{\,1/2} + B_2\,\dot{K}_2^{\,1/2} + B_{11}\dot{K}_1 + B_{22}\dot{K}_2 + B_{12}\dot{K}_1^{1/2}\dot{K}_2^{1/2}$$

B_1	B_2	B_{11}	B_{22}	B_{12}	A	Time Trend	R^2	f_1	f_2	T	s
Total											
-.44	-1.1	.051	.41	.43	-.73	.81	.81	.46	.24	-.0005	-9
(.56)	(.45)	(.26)	(.15)	(.20)	(.27)	(.26)		(.19)	(.14)	(.02)	
-.50	-1.1	.09	.44	.39	4.8		.81	.42	.24	-.002	-1.7
(.57)	(.46)	(.26)	(.17)	(.20)	(1.1)			(.19)	(.14)	(.016)	
Within											
.36	-.27	-.092	.079	.065	-1.9	.30	.10	.17	.01	-.0004	-.03
(.23)	(.21)	(.078)	(.052)	(.059)	(.44)	(.07)		(.08)	(.07)	(.0008)	
.37	-.21	-.08	.085	.042			.08	.14	.027	.0002	-.15
(.24)	(.21)	(.08)	(.052)	(.060)				(.08)	(.07)	(.0007)	
3 year averages											
1.0	-.04	.11	.37	-.18			.24	.31	.22	-.0014	-1.6
(.55)	(.50)	(.25)	(.17)	(.20)				(.19)	(.15)	(.008)	

Calculated with all functions of of $\dot{K}_2$ and $\dot{K}_2$ evaluated at the mean of the function for the data.

Note: $T = -(f_{11}f_2^2 - 2f_{12}f_1f_2 + f_{22}f_1^2)$

$$s = \frac{(f_1 f_2)(\dot{K}_1 f_1 + \dot{K}_2 f_2)}{(\dot{K}_1\dot{K}_2)T}$$

Degrees of freedom: Total 638;
Within 561;
3-year Average 115

As a final attempt to deal with both fixed effects and measurement error, the equation was estimated within, but on three year averages. This is simply a compromise between the "within" and the "between"; one hopes to smooth out the measurement error, but still have some within firm variance to control for the potential fixed effects. Not surprisingly, one gets results that lie somewhat between the other two.

There is a statistical test, discussed by Hausman (1978), and Hausman and Taylor (1981), for the presence of bias due to fixed effects. Essentially, it tests whether the "within" and "between" estimates are significantly different from each other. If no fixed effects bias is present, then both the between and within estimates are consistent, so they should not differ very much. The greater their difference, the more likely there is fixed effects bias. The Hausman test allows one to quantify this likelihood. One version of the test (Hausman and Taylor (1981)) utilizes the statistic:

$$M=(q_w - q_b)'(V_w + V_b)^{-1}(q_w-q_b)$$

where q_w is the vector of "within" parameter estimates, q_b is the vector of "between" parameter estimates, and V_w and V_b are their respective covariance matrices. Under the null hypothesis of no fixed effects bias, the statictic M is distributed as chi-squared with degrees of freedom equal to the number of coefficients.

Using the "within" estimates from Table 3-9 and analogous "between" estimates (not reported) yields a value of 310 for M; the 99% critical value for the chi-square with 5 degrees of freedom is 15. This implies strong rejection of the hypothesis of no fixed effects bias.

Unfortunately, the interpretation of this result is clouded by the presence of measurement error. This provides an alternative source of discrepancy between the "within" and "total" estimates, invalidating the test. What is not clear is whether measurement error could possibly be important enough to explain away the apparent strength of the test result.

Appendix 3-A.

Results Including Zero Patent and R&D Observations and Measurements of Scale Effects

Some of the relationships presented in Chapter 3 were re-estimated to include observations with zero patents and/or zero R&D. In addition, scale effects were also examined. The results of these exercises are presented in this Appendix.

Aggregate Patent Data

Zero Patent and R&D Observations

Some companies carry on R&D expenditures but have no patent applications, while others have patents but report no R&D expenditures. Bound, et al. (1982, p. 33) have argued that it is desirable to include such companies in the estimation; otherwise the estimates may be biased. Also, because the definition of R&D used in 10-K reporting refers to "material expenditures" for R&D, some companies, particularly smaller ones, may report zero or missing values for R&D, even though they have some level of R&D activity.

Following Bound, et al., we estimated one equation which includes firms reporting either zero or missing values for R&D and patents. This is done by setting the log of the variable equal to zero and setting to one the value of a dummy variable (PATDUM for patents, RNDDUM for R&D).

Equation 2 of Table 3A-1 shows how this procedure compares with equation 1, which includes only nonzero values. The number observations more than doubles to 999. The R&D coefficient drops to .427 (.02), which is rather close to the value of .38 reported in Bound, et al. (1982, p.34). The value of the patent dummy is -1.90, about twice the value of -.76 reported in Bound, et al.

Table 3A-1.

Log Patent Regression Results - Zero and Missing Observations Included

Variable	1 Observations = 436	2 Observations = 999
Log of R&D	.900 (.026)	.427 (.020)
Patent Dummy		-1.903 (.076)
R&D Dummy		-.025 (.088)
Other Variables Included	Intercept	Intercept
Standard Error	0.805	1.126
R^2	.7304	.6678

Figures in parentheses are estimated standard errors.

Scale Effects

Other work (e.g., Bound, et al. (1982)), suggests that firm size has a significant effect on patent behavior. These effects were investigated in two ways: by segregating the sample into small and large firms and estimating separate elasticities for each group and by estimating a quadratic form where the elasticity varies smoothly with R&D expenditures. Both approaches indicate that scale effects are important.

Table 3A-2. shows the results of dividing the sample into small and large firms. A new variable was created by multiplying log (R&D) by a dummy variable (DUML) with values zero for small firms and one for large firms. Thus the elasticities are b_1 for small firms and $b_1 + b_2$ for large firms. Two cutoffs for defining for small firms were used; sales less than $100 million (equation (2)) and sales less than $1,000 million (equation (3)). The F-statistics for both specifications were strongly significant. However, the results for equation (2) are difficult to interpret, since only eight firms in the sample have sales less than $100 million. The coefficient for log (R&D) is not significant in equation (2).

Equation (3) has significant coefficients for both variables and implies that large firms have a higher elasticity of patenting (.900 versus .735). This is consistent with results reported in Bound et al., although the latter used a quadratic specification when dividing the samples. The results for the quadratic form are reported below.

Table 3A-3 shows the results for the quadratic form. In addition, equation (3) includes the major SIC codes. The F-test for the quadratic form clearly rejects the null hypothesis in equation (2). The quadratic term is

Table 3A-2.

Log Patent Regression Results - Large and Small Firm Groups
Observations = 436

Variables	1	2*	3**
Log R&D	0.886 (0.027)	0.027 (0.142)	0.735 (0.051)
(Log R&D) * DUML	-	0.930 (0.151)	0.165 (0.047)
Other Variables	Year Dummies	Year Dummies	Year Dummies
Standard Error	0.801	0.737	0.781
R^2	0.7378	0.7593	0.7452
F-test		$F_{1,423} = 38.0$ Log R&D * DUML	$F_{1,423} = 12.2$ Log R&D * DUML

* DUML = 0 if sales < $100 million.
= 1 if sales ≥ $100 million.

** DUML = 0 if sales < $1,000 million.
= 1 if sales ≥ $1,000 million.

Figures in parentheses are estimated standard errors.

Table 3A-3.

Log Patent Regression Results - Quadratic Form

Observations = 436

Variables	1	2	3
Log R&D	0.886 (0.027)	0.664 (0.052)	0.644 (0.049)
$(\text{Log R\&D})^2$	-	0.061 (0.012)	0.039 (0.011)
Other Variables	Year Dummies, Intercept	Year Dummies	Year, 3-digit SIC Dummies
Standard Error	0.801	0.759	0.566
R^2	0.7378	0.7523	0.8123
F-test		$F_{1,423} = 24.823$ $(\text{Log R\&D})^2$	$F_{1,416} = 21.573$ 3-digit SIC Dummies

Figures in parentheses are estimated standard errors.

positive, confirming the postive scale effect. Interestingly, the quadratic form provides hardly any better estimates over simply segregating the sample into large and small firms (equation (3) in Table 3-8). The quadratic term indicates an elasticity of .94 for firms with sales of $100 million and 1.09 for firms with sales of $1,000 million.

The SIC dummies in equation (3) are still very significant as a whole ($F_{7,426}$ = 21.57). This contrasts with the results presented in Bound, *et al.* (1982, page 35), where the industry groups were not very significant in the quadratic form. It was hypothesized that the differences in average size across industry groups caused these effects to be confounded. This is evidently not the case here. However, including the industry dummies does reduce the magnitude of the quadratic coefficient by 2/3, indicating a positive relationship between scale and industry group.

The sign pattern for the industry groups is the same as in Table 3-6, except that for SIC 281, which becomes negative but with a very low confidence level. Since firms in SIC 281 tend to be the largest in the sample, this shows the effect of the correlation between scale and industry group. (On the other hand, the coefficients for SIC 281 at the 4-digit SIC level are still positive and significant.)

Line of Business Data

Zero Patent and R&D Observations

As discussed at the beginning of this appendix, the failure to include LB's with zero patents may bias the results, since with low enough R&D expenditures, there is a significant probability that a given LB will not patent in a given year, even though it does show positive patenting over a longer period of time, Bound et al (1982, p. 33). Following the method used in Bound et al., Table 3A-4 shows the effect of adding dummy variables for observations with zero patents or R&D, setting the log of the relevant variable to zero, and including all observations. The R&D coefficient does drop substantially from .655 to .432. As in the case of the firm-level regressions, this loss in magnitude is about proportional to the fraction of zero observations added to the sample. Both dummy variables are strongly significant. Following Bound et al., the coefficient of the patent dummy may be interpreted to mean that observations with zero patents have an expected value of 1/3 patent. This compares with an estimate of 1/2 patent obtained by Bound et al.

Scale Effects

The overall size of the LB unit may also affect patenting behavior. Using a cutoff value of $100 million in sales, LB's were classed as large or small, and separate intercepts and slope terms were estimated for each group. The results are shown in Table 3A-5. Equation (1) repeats the basic result from Table 3-7, while equation (2) contains the terms LDUM (= 0 for small; = 1 for large) and LDUM*Log R&D. Neither the intercept nor the slope is significant, nor is the overall F-statistic. Thus, LB size does not appear to be an important factor when taken alone.

Adding industry dummies in equation (3) does not change this result much. The F-value for the dummies is almost identical to that in Table 3-7, indicating no strong interaction between LB size and industry group. The R&D term for large LB's does become marginally significant, suggesting that the propensity to patent

Table 3A-4.

Log Patent Regression: Zero Observations Included

Observations = 399

Variables	1
Log R&D	.432 (.031)
PATDUM	-1.045 (.122)
RNDDUM	-.751 (.150)
Other Variables Included	Intercept
Standard Error	.805
R^2	.602

Figures in parentheses are estimated standard errors.

Table 3A-5.

Log Patent Regression: Effect of LB Sales Size

No. Observations = 257

Variables	1	2	3
Log R&D	.655 (.044)	.537 (.093)	.452 (.091)
LDUM		.177 (.144)	.085 (.146)
LDUM*Log R&D		.112 (.111)	.196 (.108)
Other Variables Included	Intercept	Intercept	Intercept, Industry dummies
Standard Error	.907	.902	.573
R^2	.471	.478	.782
F-test	--	$F_{2,253} = 1.63$	$F_{14,239} = 3.79$

Figure in parentheses are estimated standard errors.

increases with scale. This is consistent with the firm level results. There, however, the cutoff was $1,000 million rather than $100 million, because of the different sample size distribution.

Table 3A-6 examines the interactions between nonlinearities in R&D expenditures, sales size, and industry groups. Equation (1) repeats the basic equation, while equation (2) introduces a quadratic term for R&D expenditures. The coefficient for the quadratic term is not significant. However, adding industry dummies makes the quadratic term significant and positive. Apparently, the scale of R&D explains variation within industry groups better than overall.* This is the reverse of the results obtained with firm-level data, where the quadratic term was significant to begin with and its value was reduced by adding the industry dummies. On the other hand, the magnitudes of the linear and quadratic coefficients are quite close to those of the firm-level results when industry dummies are included. In any event, the effect is not strongly significant, and the coefficient values imply that R&D expenditures must be $24 million before the elasticity reaches unity.

Equation (4) shows the effect of adding the LB sales size intercept and slope terms. The F-test indicates the variables are not significant, but there is a clear interaction with the R&D coefficients; the standard deviation of the linear term doubles, (as was the case in Table 3-8), and the quadratic term becomes insignificant. However, this result seems due more to the weakness of the individual effects that to the strength of the interaction.

---

* This general result also occurs in the revenue function regressions.

Table 3A-6.

Log Patent Regression: Scale and Industry Effects

Observations = 257

Variables	1	2	3	4
Log R&D	.655 (.044)	.623 (.052)	.556 (.054)	.473 (.093)
$(\text{Log R\&D})^2$		.025 (.022)	.044 (.022)	.029 (.027)
LDUM				.099 (.147)
LDUM*Log R&D				.112 (.134)
Other Variables Included	Intercept	Intercept	Intercept, industry dummies	Intercept, industry dummies
Standard Error	.907	.906	.780	.781
R^2	.471	.474	.572	.575
F-test	--	$F_{1,254} = 1.25$	$F_{14,240} = 3.93$	$F_{2,238} = 0.75$

Figures in parentheses are estimated standard errors.

Chapter 4

Revenue Function*

Introduction

The second equation in the model of innovation is a revenue equation, which captures the effects of product and process innovation (in addition to capital, labor, and material inputs) on the firm's revenues. This function takes the general form:

$$Q = F(C, L, M, K, t) \qquad (4\text{-}1)$$

where:

Q = revenues, deflated by the WPI for SIC 28,
C = measure of capital stock,
L = employment,
M = materials purchased,
K = knowledge stock, and
t = time trend.

The revenue function shows how revenue varies with different mixes of inputs. The knowledge stock together with other factors contributes to total factor productivity. Total factor productivity is the amount of output produced by an average unit bundle of inputs. Treating the knowledge stock like another factor of production is equivalent to assuming that a firm can produce the same revenues with fewer normal inputs if it has a greater stock of technical knowledge upon which to rely.

As noted in Chapter 2, these increased revenues may result from greater quantity or higher price. Since only revenues (p*q) as observed, it is not possible to separate the two effects on the ouput side. However, there are measures of both product and process innovation, so their effects can be sorted-out on the input side.

---

* This is not the Revenue Function of Duality Theory, in which revenues are a function of input and output prices.

An alternative formulation of the revenue function uses value-added rather than revenues as the left-hand variable, and drops materials (M) from the right-hand side. The conditions under which either functional form produces equivalent results are identified in Fuss and McFadden (1978). Basically, the revenue function must be characterized by weak separability of the inputs. Unfortunately, neither materials nor value added are available in the firm-level data. Therefore, the revenue function is estimated with materials excluded. In the line-of-business results, the importance of this mis-specification is explored.

The specific functional form that used is the Cobb-Douglas, i.e.,

$$Q = aL^{c}C^{b}K^{d}M^{g}e^{t} \qquad (4-2)$$

with materials excluded in all but the LB results. Although the Cobb Douglas specification imposes some strong assumptions about the rates of substitution between factors, this issue is not a major concern in this analysis. Of main concern are the effects of K on Q and on any changes in this relationship which might be attributed to TSCA. In the section where the product/process distinction in investigated, K is broken down into these two components.

Greater efficiency in estimation could be achieved if it were possible to estimate the factor demand equations for labor and capital along with the revenue function. The absence of firm specific factor price data precludes this approach.

Definitions of Capital and Knowledge Stocks

Stocks of conventional capital, R&D capital, and the two types of innovation were constructed from the flow data as follows:

$$\text{Stock}_t = \text{Stock}_{t-1}(1-\text{Dep}) + \text{Flow}_{t-1} \tag{4-3}$$

where "Dep" is an assumed depreciation rate: 10% for conventional capital, and 20% for R&D and patents. An initial stock is constructed assuming that the flows have been growing at some constant rate g for a long time:

$$\text{Stock}_1 = \frac{\text{Flow}_0}{\text{Dep} + g} \tag{4-4}$$

To estimate the growth rates, the log of the flows was regressed on a time trend for each company:

$$\log(\text{FLOW}_{it}) = A + g_i t \tag{4-5}$$

This yielded estimates for each company of growth rates for physical investment, R&D, process innovation and product innovation. For conventional capital and R&D, the initial stock was created according to (4-4) based on the company's individual estimated growth rate. For the patents, these individual rates were quite erratic, because of the measurement error problem. Therefore, the average of the g_i's over all companies (.01625 for product patents and .02019 for process patents) were used instead.

For conventional capital and R&D, the initial stocks were constructed in the second year that the firm had data, using the first year's investment data. For the patent stocks, the initial stock was constructed for 1968 (or the company's first year if later than 1968). Again due to measurement error, the actual 1967 data was not used to construct the initial stock.

Instead, the predicted value from the regression of the log flow on time was used. Note that this value could only be obtained for firms having 2 or more non-zero observations for both types of patent flows. Thus product and process patent stocks were constructed for only 61 firms.

Estimates Based on Firm-Level Data

This section describes the results of the econometric analysis of the revenue function on firm-level data, ignoring the product/process innovation split. The basic equation is:

$$\log S_t = a + b \log C_t + c \log E_t + d \log K_t + e_t \qquad (4\text{-}6)$$

where:

S_t = annual sales

C_t = capital stock

E_t = employment

K_t = measure of innovation stock

e_t = error term.

The basic data set consists of 782 observations for 100 firms.

Most of the estimation work was done on a sample of firms which have positive values for inputs and sales variables and which have associated 4-digit SIC codes. (Some additional work was done including firms with zero or missing values for R&D inputs. See Appendix 4-A) The financial data was taken from firms' annual reports and 10-K reports, and covers the years 1960-1981. However, not all firms have observations for all years.

The analysis addresses the following questions: the derivation of the stocks of capital and knowledge; the lag effect of knowledge on sales; the relative contributions of across-firm and within-firm variation in explaining the behavior of sales; the effect of zero or missing observations of R&D*; the effect of the scale of R&D expenditures and the size of the firm in terms of sales*; differences in behavior across different industry groups; and the performance of patents as a measure of knowledge.

Lag Structure of Knowledge Stock

There is a significant lag time between initial R&D expenditures for a project and the eventual impact on sales which motivates that expenditure. This sequence includes basic and applied R&D steps, commercialization, and initial low volume sales. The time lag is important because it provides information about the profitability of R&D expenditures, and also it indicates the lag time that would be associated with any impacts of TSCA on the level or composition of R&D expenditures.

Table 4-1 compares three alternative specifications of the revenue function. Equation (1) contains no term for R&D expenditures. The share of capital is .201 and that of labor is .743, indicating slightly less than constant returns to scale. The R^2 is .940, indicating a high level explanatory power. Equation (2) contains R&D expenditures lagged from zero to three years and yields an unrestricted estimate of the lag structure for R&D over these years. Overall, the R&D terms are significant with an F-value of 8.58 (4,774), but they provide relatively little explanatory power; the R^2 increases only to .943. Only the last R&D term has a significant coefficient, while two others have negative but insignificant coefficients.

---

* Presented in Appendix 4-A.

Table 4-1.

Revenue Function: Effect of R&D

Observations = 782

Variables	1	2	3
Log Capital Stock	.201 (.019)	.167 (.019)	.167 (.019)
Log Employment	.743 (.022)	.665 (.026)	.660 (.026)
Year	.009 (.004)	.013 (.004)	.013 (.004)
Log R&D		-.028 (.063)	
Log $R\&D_{-1}$		.021 (.083)	
Log $R\&D_{-2}$		-.062 (.084)	
Log $R\&D_{-3}$		.189 (.068)	
Log R&D Stock			.127 (.023)
Other Variables Included	Intercept	Intercept	Intercept
Standard Error	.213	.205	.205
R^2	.940	.943	.943
F-test		$F_{4,774} = 8.58$	$F_{1,777} = 31.2$

Estimated standard errors shown in parentheses.

This implies a lag time of at least three years between R&D and significant sales impacts. The sum of the R&D elasticities is .120. Adding the R&D terms reduces the coefficients for both capital and labor. This is not surprising, since R&D has both labor and capital components.

Equation (3) includes the R&D stock constructed using the estimated growth rate method described above. The stock specification has as much explanatory power as the unrestricted lag form of equation (2) (the R^2's are the same to three decimal places), and the coefficient of .127 is significant and almost exactly the same as the sum of the R&D coefficients in equation (2). This implies that the knowledge stock formulation is a good approximation which does not sacrifice important information contained in the data.

Table 4-2 compares forms of the equation with the R&D stock lagged from zero to three years. The results are almost identical, except that the explanatory power increases slightly as the length of the lag increases. The coefficient value for the R&D stock also increases slightly, which is consistent with the finding that there is a significant lag in R&D effects. The results are too close, however, to support strongly one lag length over another. The three year lag is used in most of the subsequent regressions.

Industry Groups

The scope of the study includes 32 four-digit SIC codes in SIC 28 and 29. Separate dummy variables for each group were included in the regression, which reduces the sample size. Table 4-3 shows the results. Equation (2) shows that adding the industry dummies has significant effects on the coefficient estimates, reducing that for the capital stock by 2/3 and raising those for

Table 4-2.

Revenue Function: Alternative Lags of R&D Stock

Observations = 782

Variables	1	2	3	
Log Capital Stock	.167 (.019)	.165 (.019)	.164 (.019)	.163 (.019)
Log Employment	.660 (.026)	.659 (.026)	.660 (.025)	.660 (.025)
Year	.013 (.004)	.013 (.004)	.013 (.004)	.013 (.004)
Log R&D Stock	.127 (.023)			
Log R&D Stock$_{-1}$		.129 (.022)		
Log R&D Stock$_{-2}$			.130 (.022)	
Log R&D Stock$_{-3}$				.130 (.021)
Other Variables Included	Intercept	Intercept	Intercept	Intercept
Standard Error	.205	.205	.204	.204
R^2	.943	.943	.943	.943

Estimated standard errors shown in parentheses.

Table 4-3.

Revenue Function: Industry Groups

Observations = 716

Variables	1	2	3	4
Log Capital Stock	.095 (.019)	.031 (.021)	.039 (.021)	.041 (.024)
Log Employment	.666 (.024)	.720 (.026)	.701 (.026)	.733 (.026)
Year	.012 (.004)	.014 (.004)	.016 (.004)	.016 (.004)
Log R&D Stock$_{-3}$	.163 (.019)	.203 (023)	.189 (.029)	
(Log R&D Stock$_{-3}$)^2			-.004 (.005)	
LDUM			.273 (.076)	
Other Variables Included	Intercept	Intercept, 4-digit SIC dummies	Intercept, 4-digit SIC dummies	Intercept, 4-digit SIC dummies and slope terms
Standard Error	.165	.132	.129	.093
R^2	.944	.956	.958	.970
F-test		$F_{32,686} = 6.76$	$F_{32,684} = 6.58$	$F_{30,651} = 10.25$

Estimated standard errors shown in parentheses.

the employment and R&D stocks. The overall F-value for adding the dummies is 6.76, again not very high, in contrast to the results for the knowledge production equation.

It was hypothesized that there might be a comingling of industry and scale effects. To test this, the dummy variable for large and small firm groups and the quadratic R&D term were added in equation (3). However, this had little effect on the other coefficient estimates, suggesting that there is not a strong interaction between industry groups and size.

In equation (4), individual slope terms for the R&D stock for each SIC code are added to the separate intercepts. Overall, the slope terms are more significant than the intercepts, with an F-value of 10.25 (30,651). Fourteen out of 30 of the individual slope coefficients are significant at the 5 percent level.

Deviations from Means

Equation (1) was reestimated with all variables expressed as deviations from firm means, Table 4-4. Equation (1) is done for an unrestricted sample with 789 observations. The coefficient for the capital stock shows little change, but those for employment and the R&D stock show significant declines, and the coefficient for the R&D stock is not significant at the 5 percent level. The R^2 of .632 is rather high for "within firm" regressions.

It was suspected that the lower coefficient values resulted from insufficient variability of the variables due to the short time series available for a number of companies. To test this, the same equation was

Table 4-4.

Revenue Function: Deviations from Firm Means

Variables	1 Unrestricted Observations = 789	2 Firms with data for 1960-80 Observations = 488	3 Firms with data for 1960-80 Observations = 488
Log Capital Stock	.176 (.028)	.172 (.037)	.185 (.042)
Log Employment	.448 (.032)	.617 (.043)	.500 (.054)
Year	.020 (.002)	.016 (.003)	.006 (.003)
Log R&D Stock_{-3}	.051 (.035)	.076 (.044)	
Other Variables Included			Industry slope terms
Standard Error	.021	.018	.014
R^2	.632	.732	.804
F-test			$F_{29,455} = 5.75$

Estimated standard errors shown in parentheses.

estimated using companies with complete observations for the period 1960-1980 (22 in all). The results are shown in equation (2). The coefficients for all variables except the R&D stock are remarkably similar to those for the total regression results given in Table 4-1, equation (3), and the R^2 has risen to .732. The R&D coefficient has risen somewhat and has a higher significance level, but still is not significant at the 5 percent level. This suggests that there may not be enough non-random variability in the R&D data to get a reliable coefficient estimate.

Equation (3) shows the effects of including industry slope terms for R&D in the deviation form of the equation. Overall, the coefficients have an F-value 5.75 (29,455) which is only moderately significant. Only seven out of 21 individual coefficients are significant at the 5 percent level, and many have implausible values.

Patent Stock

In the previous regressions, R&D expenditures were used as a measure of innovative output, even though technically they measure innovative inputs. Patent applications can be regarded as a measure of knowledge production, so a knowledge stock based on patent applications was constructed to replace the R&D stock in the regression. This procedure is useful both to determine how closely different measures of innovation are related to each other and, in particular, how well patent applications perform, since they are the most accessible and broadly available measure of innovative outputs available.

Patent stocks were constructed as described above. For reasons described previously, only patent data for the years 1968-78 were used. This yielded a total number of 519 observations.

Table 4-5 shows the results for this form of the revenue function. Equation (1) includes all observations. The coefficient for the patent stock is .036 which, although significant, is a good deal lower than that for the R&D stock in various specifications. The sign of the time trend is actually negative, although not very significant. This is curious, since the general wisdom is that patents have been growing more slowly than R&D expenditures. Therefore, using patents as a measure of knowledge should leave a greater unexplained residual increase in productivity which should be attributed to the time trend. This result needs further investigation.

It was hypothesized that the low coefficient for the patent stock resulted from insufficient variability of the patent stock over time or from too much measurement error, particularly for firms with very few patents. Two approaches were tried to counteract this: restricting the sample to firms with longer time series or to firms with a minimum level of patenting activity.

Equation (2) shows the result of restricting the sample to firms with patent data for nine or more years. This has no significant effect on the coefficient estimates.

Equation (3) shows the effect of restricting the sample to firms with an average of 10 or more patents per year. This has much more marked effects on the results. The coefficient of the patent stock rises to .083 and increases in significance, although it is still a good deal below the range of coefficient values for the R&D stock (.12 to .16). These results are consistent with other work which indicates that observations with zero or low values for patent applications are noisy and difficult to use.

Table 4-5.

Revenue Function: Patent Stock

Variables	1 All Observations 1968-1978 Observations = 519	2 1968-1978 Firms with 9+ observation Observations = 474	3 1968-1978 Firms with Average Pat./Yr. = 10 Observations = 294
Log Capital Stock	.191 (.023)	.202 (.022)	.180 (.026)
Log Employment	.673 (.025)	.622 (.024)	.641 (.026)
Year	-.011 (.006)	-.004 (.005)	.003 (.005)
Log Patent Stock	.036 (.015)	.042 (.014)	.083 (.016)
Other Variables Included	Intercept	Intercept	Intercept
Standard Error	.129	.103	.044
R^2	.945	.950	.952

Estimated standard errors shown in parentheses.

In summary, one significant result of the estimation work using firm-level data is that measures of knowledge add relatively little in explaining sales. This is due largely to the high degree of correlation between R&D and other capital expenditures, and hence between the knowledge stock and the capital stock. Thus, depending on the specification of the equation, the coefficients for capital and knowledge tend to be negatively correlated, i.e. when one is high, the other low, a likely sign of high correlation between the two variables. The estimate for labor's share of total production seems more stable.

Firm and industry-specific effects have a much smaller role in explaining firm sales levels than they do for firm patenting levels. This conforms with our expectation that the patenting process is much less well understood than the production process. Finally, patents do not appear to work as well as R&D expenditures in explaining sales, but there is a significant relationship. The problem appears to arise from noise in the patent data, and is not mitigated substantially when longer time series are available.

Estimates Based on FTC Line of Business Data*

This section describes the results of the econometric analysis of the firm production function using line of business (LB) level financial data from the Federal Trade Commissions' Line of Business Survey. Two forms of the equation are estimated: one for LB sales and one for LB value-added; most of the results presented here are for the latter form. The FTC Line of Business data base covers financial data for about 441 large U.S. manufacturing corporations for the years 1974 through 1976 (1977 data now are becoming available). The data are disaggregated by individual LB code activities which correspond closely to 4-digit SIC codes.

---

* Certification: This section has been reviewed by the Manager of Federal Trade Commission's Line of Business Program for identification of individual company LB data, its release will not result in such disclosure.

Revenue Function

Table 4-6 shows some basic results for the revenue function for Line of Business data. Equation (1) includes the same kinds of variables which were used in the firm level estimation. Since the LB data includes financial measures only, payroll had to be used instead of employment. (Of course in this exercise the variable Year could only take on three values: 74, 75, and 76.) The firm-level work recognized that the equation was mis-specified because sales (rather than value-added) was the dependent variable, but materials purchased were not included on the right hand side. The availability of materials purchased in the LB data base allows us to examine the effects of this omission.

Equation (1) shows some striking differences from the similiar equation estimated for the firm data. First, the shares of capital and labor are about equal, whereas in the firm-level results, the share of labor tended to be three times that of capital. This difference is difficult to explain, and is quite robust under different specifications for each data set. The sum of the shares of capital and labor is .90, higher than that obtained in the firm-level results, but still less than 1.0. Unlike the firm-level results, the coefficient for R&D is negative and insignificant. Finally, the time trend coefficient is negative. Since the data cover the period 1974-76, this probably reflects the recessionary effect of the first oil embargo or a deterioration of the quality of the price indices during this period.

Table 4-6.

Log Sales Regression

Observations = 1046

Variables	1	2	3
Log Materials Purchased		.549 (.015)	.440 (.015)
Log Payroll	.440 (.025)	.218 (.018)	.211 (.016)
Log Plant & Equipment	.464 (.019)	.194 (.015)	.250 (.015)
Log R&D	-.003 (.013)	-.003 (.009)	.031 (.009)
Year	-.045 (.021)	-.028 (.014)	-.032 (.012)
Other Variables Included	Intercept	Intercept	Intercept, industry dummies
Standard Error	.297	.134	.093
R^2	.861	.938	.957
F test			$F_{15,1025}$ =31.73

Estimated standard errors shown in parentheses.

Equation (2) shows the results of adding materials purchased to the equation. This dramatically reduces the shares of capital and labor to a total of .41 and significantly improves the fit of the model, but has surprisingly little other impact on the estimated relations. The relative shares of capital and labor remain about equal, the coefficient of R&D stays about the same, and the time trend decreases somewhat in absolute value. These results imply that the inferences from the mis-specified firm-level equation would likely hold up if data on materials purchased could be obtained. Finally, note that including materials purchased yields a sum of coefficients for the conventional inputs of materials, labor and capital of .96, not significantly different from 1.0.

Equation (3) shows the effect of including separate intercepts for each LB industry (equivalent to 4-digit SIC code). This reduces the share of materials purchased to .44 from .55 and also causes the R&D coefficient to become significant, although the coefficient is much smaller than that obtained from the firm-level results. This smaller R&D share probably results from the shorter time series available and may also reflect economic turmoil that prevailed during that period. This interaction with the intercept dummies implies that R&D better explains differences in sales within a given industry than across industries.*

The rest of the estimation work presented here uses the value-added rather than the revenue function. It should be noted that additional forms of the revenue function were estimated, and the results were quite consistent with those of the value-added function.

---

* A similar phenomenon was found with the KPF equation.

Value-Added Equation--Basic Form

Table 4-7 shows the basic form of the value-added equation. Equation (1) excludes R&D while equation (2) includes it. The R^2 values are lower than for the revenue function, but the coefficients for labor and capital are remarkably similar to those obtained with the mis-specified production function. In equation (2), the R&D coefficient is positive and marginally significant, suggesting that R&D does better in explaining value-added than explaining sales. Its addition has little effect on the other coefficient estimates. Nevertheless, the R&D coefficient is still much lower than that obtained in the firm-level results.

Industry Effects

Table 4-8 shows the effects of including industry-specific variables in the value-added equation. Separate industry dummies are added to the basic equation (excluding R&D) in equation (2). The F-value is 32.63 (15,1026) which is highly significant and very close to the result obtained for the revenue function. This also causes the relative shares of labor and capital to change, the former decreasing and the latter increasing. This is even further at variance with the factor shares found in the firm-level results.

Equation (3) introduces R&D into the equation. As in the revenue equation, the industry dummies tend to increase the size (to .059) and the significance of the R&D coefficient.* Its F-value is 13.84 (1,1025), quite significant. Equation (4) adds industry specific slope terms for R&D expenditures. The F-value is 2.26 (15,1010), only marginally significant. Probably the short time series makes it impossible to capture industry-specific effects of variations in R&D very well.

---

* Almost identical results occurred in the revenue function.

Table 4-7.

Log Value-Added Regression--Effect of R&D

Observations = 1045

Variables	1	2
Log Payroll	.476 (.028)	.458 (.030)
Log Plant & Equipment	.427 (.028)	.420 (.023)
Log R&D		.028 (.016)
Year	-.066 (.025)	-.066 (.025)
Other Variables Included	Intercept	Intercept
Standard Error	.424	.423
R^2	.811	.811

Estimated standard errors shown in parentheses.

Table 4-8.

Log Value-Added Regression -- Industry Effects

Observations = 1045

Variables	1	2	3	4
Log Payroll	.476 (.028)	.393 (.026)	.366 (.027)	.369 (.027)
Log Plant & Equipment	.427 (.023)	.513 (.023)	.483 (.024)	.481 (.025)
Log R&D			.059 (.016)	
Year	-.066 (.025)	-.070 (.021)	-.068 (.020)	-.065 (.020)
Other Variables Included	Intercept	Intercept, industry dummies	Intercept, industry dummies	Intercept, industry dummies, industry slope terms
Standard Error	.424	.291	.288	.282
R^2	.811	.872	.874	.878
F-test		$F_{15,1026} = 32.63$	$F_{1,1025} = 13.84$	$F_{15,1010} = 2.26$

Estimated standard errors shown in parentheses.

Deviations from Means

The value-added equation was also estimated with all variables expressed as deviations from LB unit means to control for fixed effects. However, the results were unsatisfactory. The data is too noisy to identify relationships within lines of business, given that data is available for only three years.

Comparison of Firm-Level and Line of Business Results

Two other issues examined were scale effects and the effects of including observations which are zero or missing for patents on R&D. The results are presents in Appendix 4-A, and briefly summarized here. The investigation of scale effects was somewhat inconclusive. For firm-level data, there appears to be a decreasing returns to scale for large companies. When scale is measured in terms of size of R&D expenditures, there are decreasing returns to scale for R&D. For the Line of Business data, the opposite results are found. When zero or missing observations are included, the results change only a little; unlike the case of including zero observations in the knowledge production function.

As with the knowledge production equation, comparing the LB results with the firm-level results yields a mixed bag of conclusions. The single most encouraging result is that the mis-specification of the production function which results from omitting materials purchased does not seriously affect the estimate of the effect of R&D on sales or value added. This implies that firm level data, which often does not include materials purchased, may still give reliable results about R&D impacts.

A number of the discrepancies observed may be due to the short time series available for the LB data, only three years. This probably accounts for the lower coefficient for R&D obtained here, although the omission of basic R&D expenditures may also contribute to this. (This omission should be addressed in future work.)

In contrast to the results for the knowledge production equation, industry groups show more significant differences in average total factor productivity than was indicated by the firm-level data. Including industry dummies tended to add to the explanatory power of R&D; this effect was found only to a mild degree in the firm-level results. Finally, the slope coefficient for R&D was found to increase with the scale of R&D expenditures, whereas it decreased with the firm-level data. This is not easily explained and needs to be examined further.

A number of avenues could be explored further with the LB data. For example, effects of total firm R&D on growth in each LB could be examined. Effects of aggregation error due to using firm-level data rather than more detailed data could be examined precisely.

Estimates that Distinguish Between Product and Process Innovations

The Model

In this section, the revenue equation continues to include the firm's "knowledge stock" as an input, using a Cobb-Douglas production function. However, the "knowledge stock" has two components, process and product:

$$S_{it} = AL_{it}^{B_1} C_{it}^{B_2} (K_{1it} + P\,K_{2it})^{B_3} e^{B_4 t} \qquad (4-7)$$

where:

S_{it} = revenue of firm i in year t divided by a chemical industry price deflator for year t

L_{it} = number of employees

C_{it} = physical capital stock

$K_{1it} + P\, K_{2it}$ = effective total "knowledge stock"

The last term allows for an exogenous rate of productivity growth. The parameter P (which must be estimated) captures the relative value of product and process patents. It is the implicit relative shadow price of product patents to the firm, and is assumed to be the same for all firms.

This shadow price should reflect, among other factors, various institutional characteristics of the firms and their industrial setting which influence the relative value of product and process patents, including possible regulatory factors. The presence of P_t in the revenue function reflects the fact that institutional features (such as appropriability of the fruits of knowledge) as well as technological parameters influence the relationship between revenues and knowledge stocks. (For an earlier argument along these lines, see Pakes and Schankerman, (1984)).

The role of process knowledge in equation (4-7) is straightforward. Process knowledge improves productive efficiency permitting more physical output, and hence more revenue, for given capital and labor. The role of product knowledge is more indirect. For product innovation to be relevant, products must be differentiated, which implies that the firm's demand curve is downward sloping. Product innovation improves the quality of the firm's products, shifting the demand curve outward, and allowing the firm to charge a higher price for any given quantity. This also has the effect of increasing sales (p*q) without increasing capital or labor inputs; increasing productivity. If separate measures of the firm's production (q) and its price (p) were available, then the effect of product knowledge on the latter could be measured separately. However, only revenues (P*q) are observed, so these effects cannot be measured separately.

The formulation of equation (4-7) assumes that product and process knowledge are perfect substitutes in producing revenue. (The elacticity of substitution is locally infinite.) An alternative formulation would be to assume that the elasticity of substitution is unity:

$$S_{it} = A\, L_{it}^{B_1}\, C_{it}^{B_2}\, K_{1it}^{B_3}\, K_{2it}^{B_4}\, e^{B_5 t} \qquad (4\text{-}8)$$

Both equations (4-7) and (4-8) are approximations to a more complicated reality. In this case, the data seem to favor equation (4-7), so that is the formulation adopted herein.

The Results

The sample used in this section consists of 61 firms with 549 observations between 1968-78 that were complete for all variables in Equation (4-7). The distribution of observations by year is shown in Table 4-9.

The estimated equation was derived by taking logs of both sides of Equation (4-7):*

$$\log (S_{it}) = A + B_1 \log (L_{it}) + B_2 \log (C_{it}) + B_3 \log (K_{1it} + P\, K_{2it}) + B_4 t \qquad (4\text{-}9)$$

Table 4-9.

Distribution of Observations for Revenue Equation by Year

Years	Number of Firms
1968	31
1969	34
1970	40
1971	48
1972	53
1973	54
1974	55
1975	58
1976	59
1977	58
1978	59
Total	549

This equation is non-linear in the parameters and must be estimated by iterative non-linear least squares. The results of such estimation are presented in Table 4-10.

Overall, these results are quite satisfactory. The estimated elasticity of sales with respect to the total "knowledge stock" is 0.091. As will be discussed below, there is reason to believe that this is a lower bound. The shadow price of product knowledge is 0.74, significantly different from one with 99% confidence. Based on the results of these estimates, the product/process distinction is important with respect to the effect on the firm's sales. It should be emphasized here that the relative shadow price (P) is equal to the relative value multiplied by the relative propensity to patent. Thus, if firms are more likely to patent products than processes, then each product patent represents less knowledge than each process patent.

It would be desirable to estimate (4-9) in the "within" dimension, to control for firm-specific effects. Unfortunately, this is not easy in the non-linear context. Having 61 firms dummies would be quite expensive, and it is not possible to eliminate the firm effects by using deviations from firm means. This would require taking the deviations from the mean of log $(K_1 + P\ K_2)$, but this cannot be evaluated before P is known.

However, a sense of what the within regression would be like was obtained by using a linear search on P. That is, P was set to various values, the

---

* The possibility of a lag between knowledge production and its effect on sales is ignored here. As discussed above, such lags are not identifiable in this data.

Table 4-10.

Revenue Equation with Separate Product and Process Patents

Variable	
Log Employment (B_1)	.76 (.027)
Log Plant and Equipment (B_2)	.075 (.027)
Log Effective Knowledge Stock (B_3)	.091 (.017)
Shadow Price (P)	.74 (.077)
Time (B_4)	.0045 (.0055)
R^2	.94

Based on: 549 observations on 61 firms, for 1968-78. Estimated standard errors shown in parentheses.

deviations from means of all variables were calculated, and then a linear regression was estimated on these de-meaned variables. The residual sum of squares of these regressions were compared to get a rough estimate of the best P. It was found that the fit improved as P was reduced to zero, but at a very slow rate. Looking within, the value of P simply has almost no effect on the fit. Thus, if the within regression were actually estimated, this would yield an estimate of P close to zero with a very large standard error.

This result is analogous to and consistent with the KPF results using product and process patents. A precise estimate of P in the within dimension would require that the relative size of the two patent stocks vary within firms over time. As discussed in the previous chapter, this is not likely to occur. Since the variables in this equation are the stocks for the two patent types, it is particularly difficult for their relative size to vary much over time. In fact, if one looks at plots over time of the ratio of the two stocks for given firms, it is clear that it does not vary much. Thus it is not possible to eliminate fixed effects if they are present.

However, it is possible to say something about the likely direction of bias due to fixed effects. One likely source of fixed effects is variations in patenting behavior. Firms need not patent all new knowledge that they produce. Patenting is costly, and reveals information. The decision to patent involves a tradeoff between these costs and the benefit of patent protection. The terms of this tradeoff may differ for different firms. For example, small firms may have limited ability to use the innovation internally, resulting in their having a higher "propensity to patent".

The importance of this for the sales equation is that firms with a low propensity to patent have more true knowledge per patent. Thus, in the sales/patent relationship, they would have a positive residual, since their "true" knowledge is higher than that predicted by their patents. At the same time, firms with a low propensity to patent will have fewer patents, so there is a negative correlation between patents and the residual. This implies that the coefficient B_3 is biased towards zero.

As for the shadow price P, this will be unbiased, so long as firms do not differ in their relative propensities to patent product and process innovations. There may be a general difference in the propensity to patent new products versus new processes, because, for example, processes are easier to keep secret and hence less likely to be patented; and there may be firm-specific differences in the overall propensity to patent, but it seems plausible to assume that the ratio of the product patent propensity to the process patent propensity is the same for all firms. If this is the case, then P is unbiased. These propositions about the bias in B_3 and P are demonstrated algebraically in Appendix 4-B.

It is interesting to compare the results in Table 4-11 with those in Table 4-5, where the patent stock was used without distinguishing product and process patents. Allowing the shadow price of the two types of patents to differ results in an increase in the elasticity of the overall stock from about .04 to about .09. Since product patents are only "worth" 75% as much on average, mixing the two together is one source of measurement error. Eliminating this error reduces the bias and raises the coefficient.

Appendix 4-A.

Estimates Including Scale Effects
and Zero Patent and R&D Observations

Some of the functions present in Chapter 4 were reestimated to include observations that were zero or missing for patents and/or R&D. Scale effects were also examined. The results are presented in this Appendix.

Firm-Level Data

Scale Effects

The Cobb-Douglas type revenue function presented in the chapter implies that the elasticity of sales with respect to each input is independent of the levels of those variables. This is an important hypothesis to test. The effects of firm size in terms of the scale of R&D expenditures are examined here.

Table 4A-1 shows the effect of including a quadratic term for the R&D stock to capture scale effects. In equation (2), the sign of the quadratic term is negative, and has the effect of raising the linear coefficient for the R&D stock, implying decreasing returns to R&D. The effect is not terribly strong, however, with an R&D stock of $1,000 million. The coefficient of R&D drops to .150 from .205. Also, the F-value of 6.10 is not very high.

Line of Business Data

Scale Effects

The question of whether the productivity of R&D varies non-linearly with the size of R&D expenditures or LB sales was investigated. Table 4A-2 shows

Table 4A-1.
Firm-Level Data: Scale Effects
Observations = 723

Variables	1	2
Log Capital Stock	.095 (.019)	.104 (.019)
Log Employment	.666 (.024)	.661 (.024)
Year	.012 (.004)	.012 (.004)
Log RD Stock_{-3}	.163 (.019)	.205 (.026)
$(\text{Log RD Stock}_{-3})^2$		-.008 (.003)
Other Variables Included	Intercept	Intercept
Standard Error	.165	.164
R^2	.944	.945
F-test		$F_{1,717}$ = 6.10

Estimated standard errors shown in parentheses.

the results of including a quadratic term for R&D. Equation (1) repeats the basic equation from Table 4-7. Equation (2) includes quadratic the term. It is positive and quite significant, with an F value of 9.15 (1,1039). It also raises the value of the linear R&D coefficient. This result is the opposite of what was found in the firm-level analysis, namely that the quadratic term was negative and significant. There is no immediate explanation for this discrepancy. It may be related to the fact that the firm-level data included a much broader range of firm sizes, whereas the LB data basically includes only large firms. Low values of R&D may also be associated with reporting anomalies which should be investigated further.

Equation (3) shows the effect of adding industry dummies to equation (2). As before, the linear coefficient increases somewhat, but the quadratic term remains virtually unchanged. The F-value for the industry dummies is 34.23 (15,1024), very close to the other values obtained. This suggests that there is little interaction between industry type and level of R&D expenditure. It must be noted again, however, that the sample is already restricted to the largest firms.

Table 4A-3 examines the effect of classifying LB units as large or small based on a cutoff value for LB sales of $100 million. Separate intercepts and R&D slope terms are estimated for the large LB's (LDUM = 1). The two variables together are very significant in equation (2), with an F value of 77.56 (2,1038). The intercept term is highly significant, and the slope term is not significant. Adding those terms has the side-effect of making the R&D

Table 4A-2.

LB Data, Log Value Added Regression: Scale Effects
Observations = 1045

Variables	1	2	3
Log Payroll	.458 (.030)	.439 (.030)	.344 (.027)
Log Plant & Equipment	.420 (.023)	.416 (.023)	.477 (.024)
Log R&D	.028 (.016)	.053 (.018)	.089 (.018)
$(\text{Log R\&D})^2$		.013 (.004)	.014 (.004)
Year	-.066 (.025)	-.069 (.025)	-.071 (.020)
Other Variables Included	Intercept	Intercept	Intercept, industry dummies
Standard Error	.423	.420	.284
R^2	.811	.813	.875
F-test	-	$F_{1,1039} = 9.15$	$F_{15,1024} = 34.23$

Estimated standard errors shown in parentheses.

Table 4A-3.

LB Data, Log Value-Added Regression: Scale Effects
Observations = 1045

Variables	1	2	3
Log Payroll	.458 (.030)	.369 (.029)	.295 (.026)
Log Plant & Equipment	.420 (.023)	.344 (.022)	.421 (.024)
Log R&D	.028 (.016)	.006 (.019)	.034 (.177)
Year	-.066 (.025)	-.067 (.023)	-.070 (.019)
LDUM		.699 (.056)	.498 (.050)
LDUM * Log R&D		.020 (.025)	.044 (.022)
Other Variables Included	Intercept	Intercept	Intercept, industry dummies
Standard Error	.423	.369	.261
R^2	.811	.836	.885
F test	-	$F_{2,1038} = 77.56$	$F_{15,1023} = 29.45$

Estimated standard errors shown in parentheses.

coefficient insignificant and reducing substantially the coefficients for labor and capital, indicating an interaction effect. Adding intercept dummies in equation (3) does not change these results substantially.

Zero R&D Observations

Some LB's reported zero values for R&D. Because the log of zero is negative infinity, these observations were dropped in the previous regressions. To examine the effect of this omission, a modified form of the equation was estimated which included a dummy variable (RNDDUM). Observations with zero R&D had log R&D set to zero and RNDDUM set equal to one. The results are shown in Table 4A-4, equation (1) without industry dummies and equation (2) with them. The coefficient estimates are fairly similar to those in Table 4-7, equation (2) and Table 4-8, equation (3). The major difference is in the coefficients for labor and capital; here, the labor coefficient is significantly higher than the capital coefficient. The R&D coefficient is not significant in equation (1) but is significant in equation (2), and its magnitude is almost identical to that in Table 4-8, equation (3).

Table 4A-4.

LB Data, Log Value-Added Regression: Zero R&D Observations Included
Observations = 1225

Variables	1	2
Log Payroll	.536 (.027)	.483 (.025)
Log Plant & Equipment	.369 (.021)	.372 (.023)
Log R&D	.018 (.016)	.060 (.016)
Year	-.061 (.025)	-.057 (.022)
RNDDUM	.129 (.065)	-.039 (.061)
Other Variables Included	Intercept	Intercept, industry dummies
Standard Error	.527	.396
R^2	.795	.848
F test		$F_{15,1204} = 27.79$

RNDDUM = 1 if R&D = 0, otherwise RNDDUM = 0

Estimated standard errors shown in parentheses.

APPENDIX 4-B

Firm-Specific Propensity to Patent and the Revenue Function

Ignoring capital and labor, the revenue equation can be written:

$$\log S_i = B_3 \log (K_{1i} + P\, K_{2i}) + e_i \qquad \text{(B1)}$$

where e_i is a well-behaved error, uncorrelated with K_1 and K_2. We don't actually observe the K's; rather, we observe J's that are related to the K's by:

$$\begin{aligned} J_{1i} &= Q_{1i}K_{1i} \\ J_{2i} &= Q_{2i}K_{2i} \end{aligned} \qquad \text{(B2)}$$

The Q's are the propensities of different firms to patent the two forms of knowledge. To say anything definite about the bias induced by the Q's varying across firms, we must impose a little more structure. Specifically, we assume that:

$$\frac{Q_{2i}}{Q_{1i}} = c \qquad \text{for all } i \qquad \text{(B3)}$$

That is, firms differ in their overall propensity to patent, but the relative propensity to patent product and process innovations is the same for all firms. The constant c will be unity if the propensity to patent product and process innovations is the same ($Q_2=Q_1$), it will be greater than one if product innovations are more likely to be patented, and less than one if process innovations are patented more.

We now substitute the J's into the revenue equation:

$$\log S_i = B_3 \log (J_{1i}/ Q_{1i} + PJ_{2i}/ Q_{2i}) + e_i \qquad \text{(B4)}$$

multiply and divide by Q_{1i}:

$$\log S_i = B_3 \log (J_{1i} + P(Q_{1i}/Q_{2i})J_{2i}) - B_3 \log (Q_{1i}) + e_i \qquad \text{(B5)}$$

now use our assumption about $Q_{1i}/ Q2i$:

$$\log S_i = B_3 \log (J_{1i} + (P/c)J_{2i}) + [e_i - B_3 \log(Q_i)] \qquad \text{(B6)}$$

This is the equation that we actually estimated. Two points are worth noting. P is the shadow price of product knowledge. The coefficient of product patents in this equation is (P/c). That is, what we actually estimate is the shadow price of product patents, which is the shadow price of product knowledge divided by the relative propensity to patent product knowledge. If the latter is high, then each patent is worth less because it embodies less knowledge.

The bracketed term in (B6) is the error term in the equation that we actually estimate. It has two components: the well-behaved error e_i, and the "fixed effect." The latter is negatively correlated with the term in the equation that is multiplied by B_3. This is because when Q_1 is large, J_1 is large (by (B2)), and also Q_2 is large (by (B3)) and hence J_2 is large (by (B2)). Thus, for firms with big Q_1, the bracketed term in (B6) will be small and the term multiplying B_3 will be large. This implies that B_3 is biased downward.

Chapter 5

R&D Demand Function

Introduction

The third equation in the model of innovation is an R&D demand function. It closes the feedback loop by specifying that R&D expenditures (inputs to the knowledge production function) depend on the firm's expectations of future sales growth resulting from increases in the stock of knowledge via the other two equations.

The formulation of the R&D Demand Equation is based on the assumption that the firm will choose an optimal rate of R&D spending by equating the present value of the benefits of R&D to its cost. The firm's actual R&D investment in any period is then assumed to represent a "partial adjustment" to this optimum. The benefits of R&D are assumed to take the form of reductions in unit production costs, or the equivalent effect of new products. The benefits of these cost savings are assumed proportionate to total sales. The equation explicitly incorporates the specification of the knowledge production function. Under these assumptions, the optimal R&D spending is shown to depend on the firm's sales, its expected growth rate and certain parameters including the discount rate, the rate of knowledge depreciation, and the gestation lag between R&D investment and the resultant cost reduction.

Two different forms of the partial adjustment model are estimated. In the first, R&D expenditures are given as a function of last year's R&D expenditures, current sales, and the anticipated growth rate of production.

In the second, the intensity of R&D--R&D as a percentage of sales--is a function of last year's R&D intensity and the expected growth rate of production.

The model has a number of simplifying assumptions. Firms are assumed to forecast correctly (except for a random error term) the growth rate of the future sales and the benefits accruing to them in all future periods that result from current R&D efforts. Cost savings are assumed to be a constant fraction of price in each period. Appendix 5-A provides the details of the derivation of the R&D Demand Function. This Appendix draws heavily on Appendix E of the earlier Meta (1982) report.

Chapters 3 and 4 of this report provide three sets of estimation results with respect to the KPF and revenue function: 1) for a large data base of firms, 2) for models estimated with LB data, and 3) for models incorporating the product/process distinction. The R&D Demand Equation was estimated for the firm-level data base. Results were not estimated for the LB data, because it was not possible to construct suitable growth rate variables based on only three years of LB data. With respect to the product/process distinction, an adaptation of the single patent model is presented, but not empirical results.

Estimates Based on Firm-Level Data

As noted above, two forms of the R&D demand equation are estimated. They are both based on the same model of the optimal R&D stock but differ in their assumption regarding the path by which the firm moves towards this optimum. The first form of the equation corresponds to a lagged adjustment model in which R&D at the firm level adjusts over time to the desired or optimal firm level. In the second formulation, the adjustment is assumed to take place between the actual and desired level of R&D intensity, where R&D intensity is defined as the ratio of R&D to sales. As before, the models are specified in log-log form due to the Cobb-Douglas formulation of the revenue function. The form of the first equation is:

$$\ln R_t = b_0 + b_1 \ln R_{t-1} + b_2 g_t{}^E + b_3 \ln S_t + e_t \qquad (5\text{-}1)$$

where:

R_t = R&D expenditures in year t (millions of dollars),

$g_t{}^E$ = the firm's expected rate of growth of future production,

S_t = sales in year t (millions of dollars), and

e_t = error term.

In the second form of the R&D equation,

$$\ln(R_t/S_t) = b_0 + b_4 \ln(R_{t-1}/S_{t-1}) + b_2 g_t{}^E + e_t. \qquad (5\text{-}2)$$

The variable $g_t{}^E$ is not observable--no company data exist for the expected rate of growth of output. It is assumed, however, that the firm has rational expectations (Muth (1961)) regarding its growth rate. That is, its expected growth rate differs from the true growth rate by an error whose expectation is zero, and which is uncorrelated with the expected growth rate:

$$g = g^E + e, \text{ where: } E(e)=0,\ E(eg^*)=0 \qquad (5\text{-}3)$$

Under this assumption, consistent estimates can be obtained by substituting the actual realized growth rate for the expected growth rate in equations (5-1) and (5-2), and then using instrumental variables to deal with the errors-in-variables problem implied by (5-3). The actual growth rate over the subsequent 4 years is used for g, and the growth rates in each of the past 4 years as instruments.*

Ordinary least squares (OLS) regressions also were run for the sake of comparison. Because the growth rate term used in the equations is a four year average and because both current and lagged sales appear in the equations, each firm included in the data base had to have at least six consecutive years of sales and non-zero values for current and lagged R&D. This yielded a subset of 85 firms out of the larger data base of 100, for a total of 524 observations.

The Results

Table 5-1 shows the results of the R&D adjustment equations estimated in the total and within forms and with the actual (OLS) and predicted (TSLS) values of g_{avg}. The total regressions, equations (1) and (2), show an almost unitary relationship between current and lagged R&D, with the adjustment rate $(1-b_1)$ about .05. The elasticity with respect to sales is small but significant, .05 to .06 (.014). The long-run elasticity with respect to sales $(b_3/(1-b_1))$ is about 1.25 with a standard error of (.65).** The growth rate term is also significant. The coefficient for the growth terms is .25 using instrumental variables as opposed to .44 for the actual value.

---

* This differs from Meta (1982), which simply used the next year's growth rate. It was felt that R&D decisions are based on longer time horizons, and that year-to-year variation in growth rates reflects mainly noise.

** Throughout this section, standard errors of ratios of coefficients are calculated using simple Taylor expansions.

Table 5-1.

Log R&D Regression
Observations = 524

Variables	Total		Within*	
	OLS (1)	TSLS (2)	OLS (3)	TSLS (4)
Log $R\&D_{-1}$ (b_1)	.955 (.013)	.959 (.013)	.502 (.040)	.530 (.043)
Log Sales (b_3)	.057 (.014)	.051 (.014)	.362 (.051)	.218 (.099)
g_{avg} (b_2)	.440 (.142)	.250 (.176)	.828 (.211)	-.081 (.603)
Other Variables Included	Intercept	Intercept	--	--
Standard Error	.052	.053	.033	.034
R^2	.986	.986	.369	.351

Estimated standard errors shown in parentheses.

* All variables expressed as deviations from firm means.

The within results, equations (3) and (4), have an R^2 of .35 to .37. The coefficient for R&D drops to about .50; which, though still quite significant, implies an implausibly rapid adjustment. The sales coefficient rises substantially to .22 to .36, also quite significant, but the long run sales elasticity falls to about .75. The growth term is significant in equation (3) using the actual value, and not significant in equation (4) which uses the instrumental variable estimate of g_{avg}. This is because virtually all the explanatory power in the first stage regression is due to the firm's average growth rate. When we take deviations from firm means of the predicted growth rate from this first stage, there is essentially no variance left.

As shown in the derivation of the optimal R&D demand equation presented in Appendix 5-A, the coefficient on the expected growth term should be related to certain underlying parameters.

Specifically:

$$b_2 = (1-b_1)\left[\frac{1}{r+d} + T\right] \tag{5-4}$$

or

$$b = \frac{b_2}{(1-b_1)} = \frac{1}{r+d} + T \tag{5-5}$$

where:

$(1-b_1)$ = rate of adjustment towards the optimal R&D,

r = discount rate

d = rate of depreciation of knowledge

T = gestation lag between innovation and its effect on revenues

A range for the likely coefficient values can be identified by substituting plausible values in (5-5). For example:

$r = .20/\text{yr}$, $d = .30/\text{yr}$, $T = 1$ year implies $b = 3$

and $r = .10/\text{yr}$, $d = .15/\text{yr}$, $T = 2$ years implies $b = 6$

so the plausible range for b is about 3-6 years.

If b is calculated using the total TSLS estimates, then $b = .25/(1-.959) =$ 6.1 years, with a standard error of (3.44). If the within OLS estimates are used, then $b = .828/.484 = 1.71$ years, with a standard error of (2.58). Thus the estimates fall roughly in the predicted range, though they are not estimated with great precision and vary considerably depending on the specification.

Table 5-2 shows the results for the intensity form of the equation. The same pattern is seen in the adjustment coefficient (in this case $1-b_4$), with a value of about .07(.014) in the total and .55(.35) in the within. Once again the value of b (now $b_2/(1-b_4)$) can be compared to the range of 3-6 years suggested above. For the total TSLS, b is 8.3 years with a standard error of (5.1); for the OLS within, b is 3.4 years with a standard error of (2.1); and for the TSLS within b is 4.7 years with a standard error of (2.2).

A Model that Distinguishes Between Product and Process Innovations

In the first section of this chapter, the firm's optimal R&D spending was estimated as a function of its sales, its expected growth rate and the cost of new knowledge. This was in a context where R&D produced a single homogeneous stock of knowledge. An analogous equation can be derived that incorporates the product/ process distinction. However, it is rather complicated, because at any moment the firm must take account of the entire future stream of marginal revenues from both knowledge stocks. Because of this complicaton, deriving an equation for optimal R&D that is empirically useful requires a number of simplifying assumptions, such that it is not clear that the results would be useful. For this reason, an R&D demand equation incorporating the product/process distinction was not estimated. In this section, however, the derivation of such an equation is outlined, for possible use in future work.

Our starting points are the two structural equations, the KPF and the revenue equation, described in Chapters 3 and 4:

$R_t = f\ (\dot{K}_{1t},\dot{K}_{2t})$*

and:

$$S_t = A\ L_t^{B_1}\ C_t^{B_2}\ (K_{1t} + P_t\ K_{2t})^{B_3} \qquad (5\text{-}4)$$

where firm subscripts have been suppressed. The first of these equations captures the cost of new knowledge, the second the benefit. In the previous presentation, it was assumed that the firm chose an optimal level of R&D spending, in that it equated the present value of the resulting cost reduction to the cost of R&D at the margin. Now an additional dimension of optimizing behavior is added. The firm is assumed to choose the optimal mix of product and process knowledge.

A cost minimizing firm will choose its input mix by equating the relative price of any two inputs to their relative cost at the margin. In this case,

---

* Note that, as in Chapter 3, the KPF is assumed to be separable in inputs and outputs. The previous R&D demand equation derivation did not rely on this assumption. It is really not needed here either, but it does simplify the expression that results for optimal R&D.

Table 5-2.

Log (R&D/Sales) Regression
Observations = 524

Variables	Total		Within*	
	OLS (1)	TSLS (2)	OLS (3)	TSLS (4)
Log R&D/Intensity $_{-1}$ (b_4)	.930 (.014)	.934 (.014)	.463 (.035)	.407 (.036)
g_{avg} (b_2)	.683 (.155)	.546 (.193)	1.846 (.171)	2.826 (.261)
Other Variables Included	Intercept	Intercept	--	--
Standard Error	.063	.065	.036	.035
R^2	.908	.906	.439	.440

* All variables expressed as deviations from firm means.

Estimated standard errors shown in parentheses.

this means equating the shadow price of product patents -- P -- to the marginal rate of transformation of product for process patents implied by the KPF. That is:

$$\frac{f_2\ (\dot{K}_{1t},\ \dot{K}_{2t})}{f_1\ (\dot{K}_{1t},\ \dot{K}_{2t})} = P_t \qquad (5\text{-}5)$$

where the notation f_1 denotes the partial derivative of $f(\cdot)$ with respect to $\dot{K}_{1t}$.* This can be solved for $\dot{K}_{2t}$:

$$\dot{K}_{2t} = H\ (\dot{K}_{1t},\ P_t) \qquad (5\text{-}6)$$

Substituting (5-6) back into the KPF yields:

$$R_t = f\ (\dot{K}_{1t},\ H\ (\dot{K}_{1t},\ P_t)) \qquad (5\text{-}7)$$

This can be solved for $\dot{K}_{1t}$:

$$\dot{K}_{1t} = Y(R_t,\ P_t) \qquad (5\text{-}8)$$

Equation (5-8) is a reduced form for the KPF where the optimal mix of $\dot{K}_1$ and $\dot{K}_2$ has been used to eliminate one of the patent flows.

Now turn to the sales equation, and calculate the instantaneous marginal product of R&D, taking into account that the optimal K_1/K_2 mix will be chosen. That is:

$$\frac{dS_t}{dR_t} = B_3 \left[\frac{dS_t}{d(K_{1t} + PK_{2t})} \right]^{B_3 - 1} \frac{d}{dR} \ (K_{1t} + P_t K_{2t})$$

---

* Some evidence on the plausibility of the assumption that firms equate the shadow price to the MRT will be presented in Chapter 6.

$$\frac{dS_t}{dR_t} = \left[B_3 \frac{S_t}{K_{1t} + PK_{2t}}\right] \frac{d}{dR} (K_{1t} + P_t K_{2t})$$

$$= \left[B_3 \frac{S_t}{K_{1t} + PK_{2t}}\right] \left[\frac{dY}{dR_t} + P_t \left(\frac{dH}{d\dot{K}_{1t}} \frac{dY}{dR_t}\right)\right]$$

$$= \left[B_3 \frac{S_t}{K_{1t} + PK_{2t}}\right] \left[\frac{dY}{dR_t}\right] \left[1 + P_t \frac{dH}{dK_{1t}}\right] \tag{5-9}$$

using the fact that $dK_{1t}/ d\dot{K}_{1t} = 1$ by definition, in addition to equations (5-6) and (5-8).

As it stands, equation (5-9) is intractably complicated because the stocks K_{1t} and K_{2t} depend on all past flows $\dot{K}_{1t}$ and $\dot{K}_{2t}$. In order to proceed, it is assumed that stocks are simply proportional to current flows:

$$K_{1t} = A\dot{K}_{1t} \quad \text{and} \quad K_{2t} = B\dot{K}_{2t} \tag{5-10}$$

as would be the case if both stocks had been growing at constant rates for a long time, and the depreciation rates for both types of knowledge are exponential at constant rates.

Using (5-10), (5-9) can be written:

$$\frac{dS_t}{dR_t} = B_3 \; S_t \; \frac{dY}{dR_t} \left[\frac{1 + P_t dH/d\dot{K}_{1t}}{A\dot{K}_{1t} + P_t B\dot{K}_{2t}}\right] \tag{5-11}$$

In addition, it is assumed that the functions $H(\cdot)$ and $Y(\cdot)$ are homogeneous of some arbitrary degree, that is:

$$\frac{dH}{d\dot{K}_{1t}} \; \frac{\dot{K}_{1t}}{H} = n_H$$

$$\frac{dY}{dR_t} \; \frac{R_t}{Y} = n_Y \tag{5-12}$$

where n_H and n_Y are constants, not necessarily equal to unity. This allows (5-11) to be rewritten as:

$$\frac{dS_t}{dR_t} = \frac{B_3 S_t n_Y}{R_t} \quad \frac{1 + P_t n_H \, (\dot{K}_{2t}/\dot{K}_{1t})}{A + B\, P_t \, (\dot{K}_{2t}/\dot{K}_{1t})} \tag{5-13}$$

Note that, because of equation (5-5), the ratio $\dot{K}_{2t}/\dot{K}_{1t}$ depends only on P_t.* Writing $P_t \, (\dot{K}_{2t}/\dot{K}_{1t}) = g(P_t)$, equation (5-13) reduces to:

$$\frac{dS_t}{dR_t} = \frac{B_3 S_t n_Y}{R_t} \quad \frac{1 + n_H \, g\,(P_t)}{A + B\, g\,(P_t)} \tag{5-14}$$

The optimal R_t is that which equates the discounted stream of these instantaneous marginal products to the cost of R&D; since R&D is expressed in dollars, the latter is 1. Therefore, R_t^*, the optimal level of R&D, is given by the solution of:

$$1 = \int_{t+T}^{\infty} \frac{dS_u}{dR_t} \, e^{-r(u-t)} \, du \tag{5-15}$$

where T is the mean lag between R&D expenditure and its effect on sales. Making the additional assumption that both forms of knowledge depreciate at the same rate d this becomes:

$$R_t^* = B_3 \, n_Y \int_{t+T}^{\infty} S_u \, \frac{1 + n_H \, g\,(P_u)}{A + B\, g\,(P_u)} \, e^{-(r + d)(u-t)+dT} \, du \tag{5-16}$$

If it is assumed that sales will grow at the constant rate g, such that:

$$S_u = S_t e^{g(u-t)}$$

---

* This is where separability helps. Without it, the function g(.) in (5-14) would be a function of P_t *and* R.

then this becomes:

$$R_t^* = e^{dT} \quad B_3 \, n_y \, S_t \int_{t+T}^{\infty} \frac{1 + n_H \, g \, (P_u)}{A + B \, g \, (P_u)} \; e^{(g-(r+d))(u-t)} du \qquad (5\text{-}17)$$

Equation (5-17) gives an expression for the optimal R&D at time t, but it depends on the entire future stream of shadow prices, $[P_u]$, in a complicated way. To arrive at an estimable form, further assumptions would be needed. One route would be to assume that A = B, i.e. product and process patents have the same growth rates, and that $n_H = 1$. In this case (5-17) would reduce to a form quite similar to that estimated previously in the context of a single aggregate knowledge stock. In that formulation, it is not possible to say anything about the effect on R&D demand of changes in the relative cost or productivity of product versus process innovation.

An alternative simplification, which preserves the product/process distinction, is that firms hold static expectations about P_u. That is, firms choose R_t as if they believed that $P_u = P_t$ for all u t. Given the institutional features that determine P, this does not seem implausible, and it is a common assumption in the investment literature. In this case, (5-17) simplifies to:

$$R_t^* = \left[\frac{B_3 n_y S_t}{r + d - g}\right] \left[\frac{1 + n_H \, g \, (P_t)}{A + B \, g \, (P_t)}\right] e^{(g-r)T} \qquad (5\text{-}18)$$

This gives a (non-linear) equation for the optimal R&D spending as a function of A, B, S, g, P and certain parameters. These parameters include those from before (r, d and T) and three that are parameters of the KPF (n_y

and n_H) or the revenue function (B_3). To derive an estimating equation, take a linear approximation, and then add a partial adjustment mechanism as done previously. The independent variables, in addition to S and g, would be A and B, which can be evaluated for each firm based on their product and process patent time series, and the shadow price P. While P is not observed, an estimate of P can be calculated for each firm in the form of the MRT for the firm calculated from the KPF parameter estimates. Under the assumption that this MRT can proxy for the firm's shadow price, an R&D investment equation can be estimated, based on (5-18).

Appendix 5-A

Derivation of the R+D Equation*

The equation describing the determinants of R+D is derived from relations that govern the effects of R+D expenditures on costs. As is implicit in the production function, we assume that the average costs to the firm vary directly with the stock of knowledge, i.e., that a measure of knowledge exists such that average costs, C, are given by

$$C = \frac{c(w)}{K}$$

where

$c(w)$ = average variable (non-R+D) costs,
w = factor prices, and
K = the stock of knowledge.

Increments to knowledge, $\dot{K}$ are produced by R+D according to $\dot{K} = \bar{R}^{\alpha}$, where $\bar{R}$ is the stock of R+D.

The stock of knowledge at time t is the sum of all prior increments to knowledge discounted to reflect their loss in productive value over time. Thus,

$$K_t = \int^{t} \dot{K}_s \, e^{-\lambda(t-s)} ds$$

where λ is the rate of obsolescence, and t and s are time subscripts. Assuming the growth rate of research to be constant prior to the sample period, say g_r, the stock of productive knowledge becomes:

---

*This derivation is due to Mark Schankerman.

$$K_t = \frac{R_t^{\alpha}}{\lambda + \alpha g_r}$$

These relations can be used to find the marginal savings in costs due to research expenditures:

$$\frac{dC_t}{dR_t} = \frac{-\,c(w)\ dK_t/dR_t}{K_t^2}$$

Since $dK_t/dR_t = d\dot{K}_t/dR_t$ we can write

$$\frac{dC_t}{dR_t} = -\left(\frac{c(w)}{K_t}\right)\frac{d\ln K_t}{dR_t} = -\ \frac{C_t}{R_t}\ \alpha$$

where λ is the rate of obsolescence and t and s are time subscripts.

The expected benefits to R+D expenditures at time t are

$$B_t = -\frac{dC}{dR}\,Q_t\,e^{-\delta t} = -\frac{dC}{dR}\,Q_0 e^{(g-\delta)t}$$

where

Q_t = production rate at time t,
δ = rate of decline in the value of benefits due to competitive forces, and
g = expected growth rate of production.

Thus

$$B_t = \alpha\ \frac{C_t}{R_t}\ Q_t\ e^{(g-\delta)t}$$

Under the assumption that a firm will spend on R+D up to the point at which the marginal dollar yields a dollar in the present value of the benefits,

$$1 = \int_T^{\infty} B_t\, e^{-rt}\, dt$$

where

r = discount rate and
T = lag between R+D expenditures and the generation of benefits.

Thus

$$1 = \int_T^\infty \alpha\, k \frac{C_0}{R_0} Q_0\, e^{(g-r-\delta)t} dt$$

which (if $r + \delta \geq g$) can be integrated to yield

$$1 = \alpha\, k \frac{C_0 Q_0}{R_0^*} \quad \frac{e^{(g-r-\delta)T}}{r + \delta - g}$$

where R_0^* is the optimal level of R+D expenditures.

Now we solve for R_0^* and take the log of the result to get

$$\ln R_0^* = a + \ln C_0 Q_0 - \ln(r+\delta-g)$$

where

$$a = \ln \alpha + (g-r-\delta)T$$

By the first order Taylor's expansion around $g = 0$,

$$\ln(r+\delta-g) = \ln(r + \delta) - \frac{1}{r + \delta} g$$

Substituting,

$$\ln R_0^* = a' + bg + \ln C_0 Q_0$$

where a' and b are constants:

$a' = \ln[\alpha/(r + \delta)] - (r + \delta)T$
$b = 1/(r + \delta) + T$

Let P_0 be the price of output. Then

$$P_0 = C_0 + Z$$

where Z is the payment to R+D inputs equal to the currently accruing quasi-rents. If we assume that Z is equal to a constant proportion of R+D costs, $\emptyset$, then

$$P_0 = C_0 (1 + \emptyset)$$

where $\emptyset = Z/C_0$. Then

$$\ln P_0 = \ln C_0 + \ln(1 + \emptyset)$$

and

$$\ln C_0 = \ln P_0 - \ln(1 + \emptyset)$$

Substituting in the R+D equation gives

$$\ln R_0^* = a'' + bg + \ln P_0Q_0$$

where g is the expected rate of growth of production and P_0Q_0 is current sales.

The above equation represents the desired rate of R+D expenditures in a static equilibrium. We introduce a partial adjustment model as follows:

$$\ln R_t - \ln R_{t-1} = \gamma (\ln R_t^* - \ln R_{t-1})$$

where γ is a constant adjustment rate. The resulting equation for R+D demand is

$$\ln R_t = (1 - \gamma) \ln R_{t-1} + \gamma \ln R_t^*$$

$$= (1 - \gamma) \ln R_{t-1} + \gamma a'' + \gamma bg + \gamma \ln P_tQ_t$$

or

$$\ln R_t = \beta_0 + \beta_1 \ln R_{t-1} + \beta_2 g + \beta_3 \ln P_t Q_t$$

where the β_i are constants as follows:

$$\beta_0 = \gamma a'' = \gamma[a' - \ln(1 + \phi)]$$

$$= \gamma\, [\ln(\alpha/(r + \delta))] - (r + \delta)T - \ln(1 + \phi)]$$

$$\beta_1 = 1 - \gamma$$

$$\beta_2 = \gamma b = \gamma\left(\frac{1}{r + \delta} + T\right)$$

$$\beta_3 = \gamma$$

The average lag between R+D expenditures and the returns to R+D is*

$$L = \frac{1 - \gamma}{\gamma} + T$$

An alternate model with the adjustment equation in terms of R+D intensity (R+D as a percent of sales) is expressed as follows:

$$\ln R_t - \ln P_t Q_t = \beta_0 + \beta_4\, (\ln R_{t-1} - \ln P_{t-1} Q_{t-1}) + \beta_2 g$$

---

*It is well known that the specification associated with partial adjustment models is identical to the problem of estimating a Koyck geometric lag of the form

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} w^i X_{t-i} + \epsilon_t$$

where the adjustment factor γ is equivalent to the lag factor w. The mean lag is the average lag period weighted by the individual lag coefficients, and it is straightforward to show that this is equal to $(1-\gamma)/\gamma$. In addition, the constant gestation lag θ must be accounted for, yielding the result in the text. See any econometrics text, e.g., Pindyck and Rubinfield (1976), pp. 213-215.

Chapter 6

Implications of the Results, Extensions of the Model, and Its Use for TSCA Impact Analysis

Introduction

In this report we have developed a model of the innovation process in the chemical industry, collected a large set of data at the firm level, and applied the model empirically to these data. The model of the innovation process consists of three parts: 1) a revenue function which links the revenues earned by a firm to its quantities of capital and labor, and a measure of either its innovative input (R&D expenditures) or an indicator of its innovative output (patents); 2) an equation linking the R&D input of a firm to a measure of its R&D output, patents; and 3) an equation which determines the profit maximizing level of R&D effort which a firm chooses to undertake. The model performs quite well empirically, in terms of both its overall explanatory power and the various predictions of specific coefficients generated by the model. Since the empirical results of the model are described in detail in the preceding three chapters, we will not attempt to summarize them here. Instead, we begin by interpreting the main findings in terms of the implied marginal products and rates of return to the various inputs of interest (capital, R&D, and patents in the KPF and the revenue functions). We then suggest some main lines of further empirical research which can be undertaken within the framework developed in this report and with the available data. Finally, there is a discussion of how the empirical results and the model can be used as a framework to assess and simulate the impacts of TSCA on innovation in the chemical industry.

Interpretation of Results

The estimates of the patent production function and the revenue function allow us to estimate marginal products of R&D in producing patents and sales, and of patents and capital in generating sales. The log linear form of the patent production and revenue functions yield direct estimates of the elasticities, which must be converted to marginal products. For example, the marginal product of the R&D stock in generating patent flows is equal to its elasticity times the ratio of the patent flow to the R&D stock. Since this ratio varies over the sample, we choose to evaluate this (and other) marginal products at the sample mean of the patent-R&D stock ratio.

Table 6-1 summarizes the marginal product calculations. Turning first to the patent-R&D relation, using the elasticity estimate of .969 (standard error of .033) from Table 3-3, equation 3, and the mean patent/R&D value of .493, we obtain an estimate of .48 (.016). This says that an increase in the R&D stock of one million 1972 dollars would produce approximately one half of an additional patent during the first year. Since a one million dollar expenditure on R&D flow today augments the R&D stock today by the same amount, we can also interpret this result as saying that the contemporaneous impact of that increase in R&D flow is an additional one half of a patent. However, the additional R&D flow also augments future values of the stock of R&D, with the amount declining at the depreciation rate appropriate for the computation of the stock of R&D in the patent production function. If, as assumed in the stock construction, the depreciation rate for R&D is 20 percent, then the additional one million dollar R&D investment eventually produces a total of

Table 6-1.

Estimates of Marginal Products

A. Patent Production Function	
1. R&D Stock	
Mean of Patent/R&D stock (millions $)	0.493
Elasticity	0.969
Short-term marginal product (patents)	0.48
Long-term marginal product (patents), assuming 20% depreciation	2.50
B. Revenue Function	
1. Capital Stock	
Mean of Revenues/Capital stock (millions $)	2.55
Elasticity	0.167
Marginal product (millions $)	0.43
2. R&D Stock	
Mean of Revenues/R&D stock (millions $)	9.34
Elasticity	0.127
Marginal product (millions $)	1.12
3. Patent Stock	
Mean of Revenues/Patent stock (millions $)	9.51
Elasticity	0.036
Marginal product (millions $)	0.34

about 2.5 patents (.5/.2 = 2.5). Since R&D is the only factor of production generating patents in this scheme, the marginal cost of a patent is the reciprocal of the marginal product of R&D. Therefore, we would estimate the marginal cost of an additional patent at about $400,000.

Turning to the revenue equation, we can similarly evaluate the marginal products of the capital and R&D stocks (using Table 4-2, equation 1) and of the patent stock (using Table 4-5, equation 1). At sample means, a one million dollar increase in the capital stock increase revenues by .167 x 2.55 = .43 (standard error of 0.48) million dollars. A one million dollar increase in the R&D stock raises .127 x 9.34 = 1.12 (.21) million dollars of revenue, while an increase in the patent stock of one patent generates .036 x 9.51 = .34 (.14) million dollars. These are the contemporaneous effects, and they can be converted to long-run streams of benefits using appropriate depreciation rates. Rather than do that here, we now turn to the measurement of the implied rates of return to investments in these assets and ask, what implicit rate of discount would make the present value of the sales generated by such an investment just equal to the cost of the investment? One caveat is necessary first. We are interested actually in the marginal products and rates of return to these investments in terms of value-added. Because we have neglected the materials input in the estimation of the revenue function, our estimated marginal products overstate the return to these inputs. As a reasonable approximation, we assume that the cost of materials is half of sales, and hence the marginal products in terms of value-added are about half of those reported in Table 6-1 and in the text above.

Pakes and Schankerman (1984) show that the internal rate of return to a dollar investment in a stock of an asset, denoted by r, can be written implicity as: $e^{rT}(r+\delta) = MP$, where MP is the marginal product of the stock, δ is the appropriate depreciation rate, and T is the gestation lag between the investment and the beginning of the resultant stream of revenues. If $T = 0$, then this reduces to $r+\delta = MP$, or the gross rate of return is the marginal product of the stock. Consider first the case of traditional capital. Using the estimated marginal product (adjusted for materials) and assuming that $T = 0$, we get a gross rate of return before taxes of about 22 percent. If we assume an average tax rate of .35 and a depreciation rate for capital of ten percent, this yields an implied net after-tax internal rate of return of about 4 percent. Next consider the case of R&D. Using the marginal product estimate adjusted for taxes and materials, and assuming that $T = 2$ (years) and a depreciation rate of 20 percent (which is probably low, see Pakes and Schankerman), we get a net after-tax internal rate of return of about 9 percent. However, a caveat is required here. In the case of R&D, what we measure is approximately the excess return, above and beyond the return to traditional factors of production because the R&D inputs are also included in the measurement of the other factors of production in the revenue equation (Schankerman, 1981). In other words, if the rate of return to R&D were the same as that to traditional inputs, then the coefficient of R&D would be zero. Since it is not, it indicates the amount by which the rate of return to R&D inputs exceeds the return to conventional inputs.

Finally, we can compute the rate of return to patents. Here some modification of the preceding rate of return formula is needed, since we are considering a unit investment in patents rather than a dollar investment in patents. The correct formula is (setting $T = 0$): $r + \quad = MP/MC$ where MP is the marginal product of the stock of patents and MC is the marginal cost of a patent. Using the estimated marginal product adjusted for taxes and materials, the estimated marginal cost of $400,000, and again assuming a gestation lag of two years and a depreciation rate of 20 percent yields an after-tax net rate of return to patents of about 6 percent. Again, since the R&D that produced the patents is included in measured capital and labor, this is an excess return. It is slightly lower than our estimate of the excess return to R&D. This is probably because the underlying elasticity estimate is biased downwards due to patents being an error-ridden measure of knowledge production.

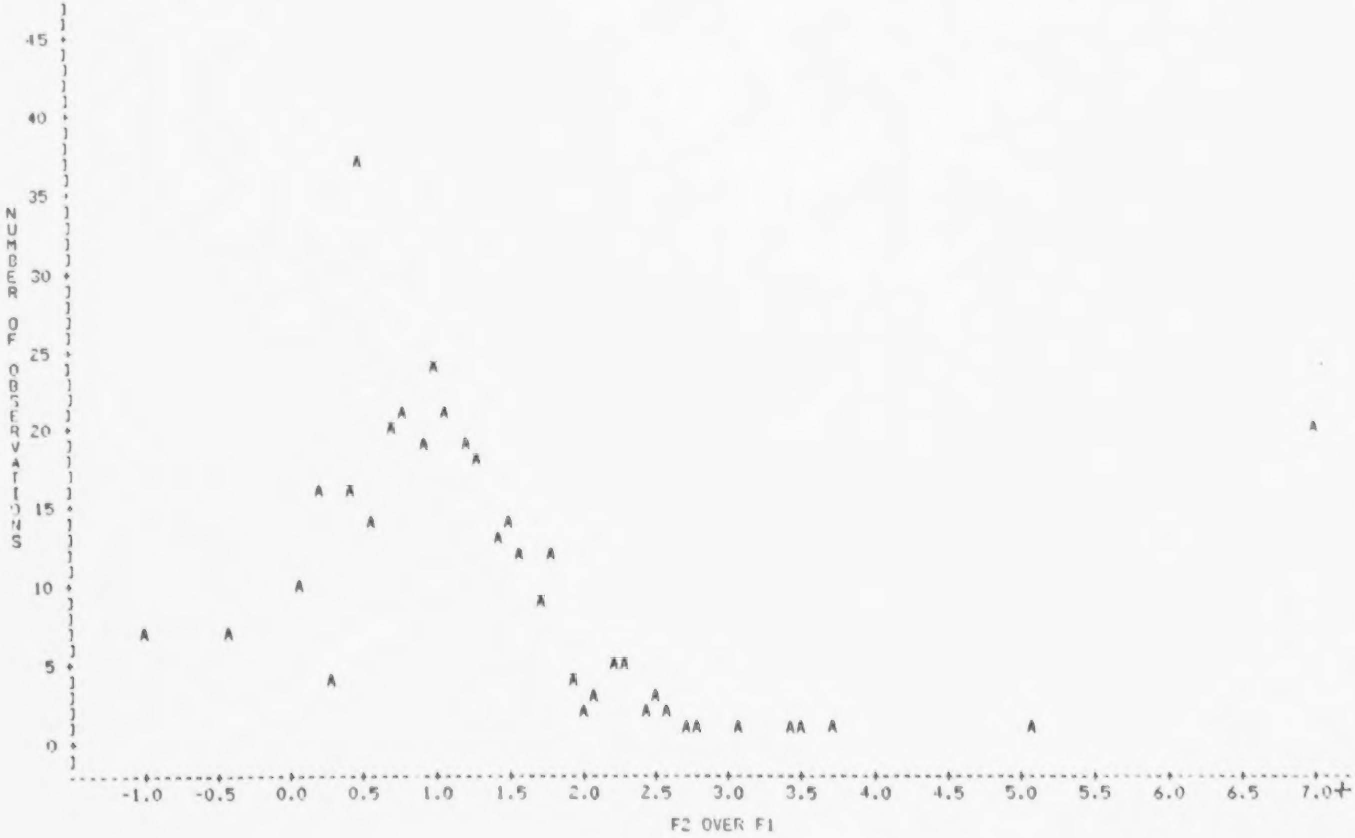
These computations of the after-tax rates of return to the capital assets in the model - capital, R&D, and patents - are very reasonable. They are broadly similar across assets, as would be expected if the firms are profit maximizing in their choice of these inputs, and they are quite consistent with the relevant economic literature. These estimates reinforce our belief that the empirical results in the earlier chapters are reasonable and that the empirical implementation of the model is quite successful.

Main Lines for Further Research

We now briefly outline some possible and interesting extensions of the empirical work and the model which could be undertaken. First, in the present report the three basic relationships of the model (patent production function, R&D demand, and revenue function) are estimated separately, even though they are conceptually linked. A useful alternative strategy would be to recognize the elements of simultaneity in the model and structure the estimation procedure accordingly. For example, simultanous estimation of the system of equations would recognize that the expected rate of growth of demand on the right hand side of the R&D demand equation must be an endogenous variable since it is the outcome of past R&D decisions via the revenue function. The main purpose of this exercise would be to purge the empirical estimates of whatever simultaneous equations bias may be present.

A second, and we think quite interesting, extension concerns a behavioral linkage between the revenue function and the knowledge (or patent) production function in the presence of the product-process patent distinction. As noted in the derivation of the optimal R&D demand equation in Chapter 5, a profit maximizing firm should equate the shadow price of product patents -- which is a parameter of the revenue equation -- to the marginal rate of transformation (MRT) between product and process patents, which can be calculated from the parameters of the knowledge production function and the data. At present, this linkage is not embodied in the estimation of the model. Figure 6-1 shows a plot of the frequency distribution of the estimated MRT in the data, based on the estimates of the KPF presented in the first row of Table 3-9. Recall that our estimate of the shadow price (P_t) from the revenue function with two types of patents is 0.74. We can roughly interpret this as an average of the actual shadow prices which differ across firms. The distribution in Figure 6-1 peaks at about 0.7, and the mass of the distribution is in the range 0.25-1.5. This rather remarkable finding seems to be roughly consistent with the revenue equation estimate under the assumption of optimizing behavior. This suggests the possibility that in future work the KPF and the

FIGURE 6-1. Distribution of the Marginal Rates of Transformation



revenue equation could be estimated jointly, imposing (and testing statistically) the behaviorial assumption that firms choose the optimal mix of product and process patents, at least on the average over time.

A related extension involves a fuller modelling effort on the determinants of the relative shadow price of product patents which a firm faces. At present we estimate this parameter in the revenue function, constrained to be equal for all firms. As noted in the preceeding suggestion, we could proxy this parameter by the estimated marginal rate of transformation from the KPF, which of course permits it to vary across firms. Ultimately, though, one would want to try to explain, or at least to find empirical correlates for, variations in this shadow price across firms. Is it related to the size of the firm, the industrial composition of the firm's output, or other variables? By identifying such variables and estimating their effect on the choice of a firm's patent mix, we could then be in a better position to assess the impact of TSCA on the patent mix and hence on a firm's revenues.

The fourth extension is to estimate the R&D demand equation which distinguishes between product and process patents. This estimation effort requires a number of simplifying assumptions, and also depends on paramters estimated from the KPF and revenue functions, as noted in Chapter 5.

One final extension of the model, a rather substantial one, would be to incorporate in the model an equation relating the stock market value of the firm to other variables in the model, in particular to R&D and patents. At present, the impact of innovative input or output (R&D and patents) on the economic fortunes of the firm is captured by the revenue function. It would be useful to go one step further and estimate the impact of innovation activity on the profitability of the firm. The advantage of using an equation for the stock market value of the firm is that, unlike contemporaneous measures of profits, the value of the firm should reflect expectations of future profit streams due to current innovative activity. Extensions of the model in this direction would enhance our ability to assess the impact of TSCA regulations on the profitability of firms, via their effects on the innovation activities of these firms.

Finally, there are more things to be done with the Line of Business data. In a number of ways the empirical work with these data proved unsatisfactory. This is particularly troubling, since these data should avoid the problem of aggregating diverse activities of a firm into a single line. Part of the problem may be that the reporting of these numbers is subject to large errors. The "output" of the R&D program of a line of business may be fungible in that a new product or process may be useful to many lines of business within a firm. An alternative use of these data to achieve a different level of aggregation may be more productive. Instead of using each line of business separately, they could be used to distinguish the chemical from the non-chemical activities of the firm, both in terms of revenues and expenditures. Then the chemical lines of business could be aggregated to enable us to study just the chemical-related portion of each firm's business. Another alternative is to maintain the lines of business distinction and recognize that we then have several "cross sectional" dimensions to the data (in addition to the short time series dimension). In particular, we would have the identification of the line of business, the firm, and the 3-digit industry of the firm. Such a multiple dimension structure to the data may enable us to treat in a more serious way the problems of errors in the variables and fixed firm effects (see Pakes (1983) for the methodology).

Use of Model to Assess Impacts of TSCA

There are several ways in which the model can be used to assess the economic impacts of TSCA. One approach would be to reestimate the same equations on post-TSCA data to determine if the relationships have been altered. However, this approach requires a determination of whether or not these changes are in fact due to TSCA or to other events, and it requires sufficient post-TSCA data.

Alternatively, and we think more operationally, the impacts of TSCA can be assessed using the relationships as estimated in Chapters 3 through 5. The model developed in this report relates "endogenous" variables of policy interest, such as a firm's revenues, R&D activity and patents, to other "exogenous" variables such as the expected growth of sales, shadow price of patents, gestation lags, and so on. Some of these "exogenous" variables, or levers, in the model may be influenced by the operation of TSCA regulations. Once the impacts of TSCA on these levers are identified, the effect of such changes on the "endogenous" variables of interest can be evaluated by simulation exercises.

The first step in any such analysis must be to gather information, or informed guesses, as to how TSCA will likely affect the "exogenous" variables in the model. The magnitude of the changes depends on the administration of the PMN process and on company reactions to TSCA requirements. For example, actual data on the number of PMN's filed, the length of time between filing and EPA's ruling, the amount of additional information required from the company and its associated cost, and the disposition (including withdrawal by the company) can be used to adjust the levers in the equations of the model to represent the impacts of TSCA. This sort of information, or some informed guesses, are indispensable to assessing the impacts of TSCA. This is true for any model of the innovation process, not just the one developed in this report. This type of information has not been gathered by the authors of the present report, and it may be most effectively provided by members of the EPA in conjunction with industry specialists.

For reference, we summarize the basic model with the distinction between product and process patents as follows:

$$(1) \qquad S = f(\overset{+}{K}, \overset{+}{L}, \overset{+}{P_1}, \overset{+}{P_2}, \overset{+}{p})$$

$$(2) \qquad R = g(\overset{+}{\dot{P}_1}, \overset{+}{\dot{P}_2})$$

$$(3) \qquad R = h(\overset{+}{g_s}; \overset{-}{T}, \overset{-}{\delta})$$

Equation (1) is the revenue function relating sales to the quantities of capital and labor, and to the stocks of product patents P_1 and process patents P_2. The relative shadow price of product patents, p, parameterizes this equation. Equation (2) is the KPF relating the flow of product and process patents, $\dot{P}_1$ and $\dot{P}_2$, to the R&D input. Recall that if firms choose their patent mix so as to maximize profits, then the marginal rate of transformation between $\dot{P}_1$ and $\dot{P}_2$ is equated to the shadow price, p. This yields the ratio of $\dot{P}_1/\dot{P}_2$ as a function of p, say $\dot{P}_1/\dot{P}_2 = H(p)$, and the parameters of this function are the same as those appearing in the KPF, equation (2). Finally, equation (3) is the R&D demand equation relating the expenditures on R&D to the expected growth in sales, g_s, the gestation lag T, and the rate of obsolescence in the market δ. For easy reference, we have noted the qualitative effect of each variable in the functions (1-3) on the dependent variable by a positive or negative sign. Also recall that the parameters of the KPF appear in the R&D demand function (see Chapter 5).

Given this setup, the principal levers in the model are the following:

1. parameters of the KPF, that determine the marginal cost of product and/or process patents (at a given mix of patents),
2. expected growth rate of product demand ("g_s" in the R&D demand equation),
3. gestation lag ("T" in the R&D demand equation),
4. depreciation rate of patents or knowledge ("δ" in the R&D demand equation), and
5. shadow price of product patents.

With reference to the model summarized in equations (1-3), we now briefly identify the role of each of these levers on the variables of interest. First TSCA regulations may raise the cost (in terms of R&D) of producing patents, particularly product patents since new chemical products are the specific focus of those regulations. The marginal cost of patents is derived directly from the KPF, equation (2), and depends on the parameters of that equation. Information on how the cost will be raised can be translated directly into shifts of the parameters of the KPF. This leads to two effects: (1) given the shadow price, the optimal mix of $\dot{P}_1$ and $\dot{P}_2$ will change, since the marginal rate of transformation between $\dot{P}_1$ and $\dot{P}_2$ is a function of the KPF parameters (and $\dot{P}_1/\dot{P}_2 = H(p)$); (2) since the KPF parameters enter the R&D demand equation, the firm's choice of its R&D spending will change.

The second possible impact is on the expected growth rate of sales. If TSCA results in a reduction of this expected growth rate, this will lead directly to a reduction in the R&D spending via equation (3). It may seem unlikely that TSCA would affect the expected growth of sales directly, but such an effect may occur indirectly. For example, as we suggested, if TSCA raises the marginal cost of producing patents, one of the consequences would be a decline in R&D spending. This would yield a smaller number of patents produced via equation (2), and hence a lower rate of growth of sales via the revenue function (1). If firms anticipate this change and base their expectations of the growth of sales on this understanding (as they would under the "rational expectations" formulation used in this report), then they will revise downward their forecast of g_s.

Third, it is likely that TSCA will increase the gestation lag, the time between an expenditure on R&D and the generation of initial benefits. The impact on the gestation lag may vary across firms depending on the actual length of the PMN review process and the point in the development process at which the firm files its PMN. The gestation lag ("T") is a central feature of the R&D demand function. We assume that a firm will increase its R&D expenditures up to the point at which the marginal dollar yields a dollar in present value of the benefits. If the gestation lag rises, then the present value of the monetary benefits received for a dollar of R&D decreases and the optimal level of R&D declines, as indicated in equation (3). The effects of a rise in the gestation lag can be analyzed directly from (3)*.

The fourth lever is the private rate of obsolescence faced by firms in the market, δ. If TSCA does result in a reduced rate of innovation, this effect will be partially offset by the effect of the private rate of depreciation on the incentive to do R&D. Reduced innovation, or at least the introduction of fewer new products on the market, may lead to a decline in the effective private rate of depreciation. Such a fall in the value of will result in greater benefits to the firm from a given R&D expenditure level. Since firms are assumed to increase R&D until the marginal dollar produces benefits equal to the cost at the margin, a decline in results in an increase in the level of optimal R&D. This appears explicitly from the role of the parameter in the R&D demand equation, equation (3).

The final consideration involves the relative shadow price of product patents, p. Since TSCA applies more directly to product innovation, it may reduce the relative value of product innovation, that is, reduce the shadow price. The shadow price plays two roles in the model (1-3). First, it appears as a parameter in the revenue function as a relative weight on the stocks of the two types of patents. And, second, it determines the optimal mix of product and process patents in the KPF for a profit maximizing firm. Therefore, changes in the shadow price would have two direct effects. First, for a given mix of product and process patents, changes in p would lead to changes in the revenues of a firm, via the revenue function. Second, a reduction in p would induce optimizing firms to shift their R&D toward process innovation, in a precise way governed by the relation $\dot{P}_1/\dot{P}_2 = H(p)$ which we derive from the KPF. This in turn would have the indirect effect of changing the mix of patent stocks $\dot{P}_1$ and $\dot{P}_2$, and hence affecting revenues, via the revenue function (1). Finally, as in the case of anything which changes revenues, the realized growth rate of sales will be affected. This in turn will affect the expected growth rate of sales and hence the optimal level of R&D via equation (3).

---

* One point is required here both for the role of the gestation lag and the rate of depreciation in the R&D demand equation. As reference to the derivation of the equation in Appendix 6, P. 4 shows, the optimal level of R&D varies inversely with both T and δ. To get estimable forms, however, we then take Taylor approximation to the nonlinear pieces in the equation and these parameters then appear in two places, with countervailing influences, in the estimated equations. The theoretical role of the parameters is unambigous, however, as the derivation will confirm.

This concludes the overview of how the model can be used to simulate the effects of TSCA regulations on the variables of policy interest. We emphasize again that any such analysis first requires that one gather information about how TSCA will affect the various levers in the model. At that point the model can be enlisted to trace through, in an internally consistent way, the effects of these changes on the innovative performance and economic fortunes of the firms in the chemical industry which are subject to these regulations.

References

Allen, R.G.D. (1938) *Mathematical Analysis for Economists*, p. 342.

Bound J., *et al*., (1982) "Who Does R&D and Who Patents " National Bureau of Economic Research, Working Paper No. 908, Cambridge, MA.

Diewart, W. E. (1974) "Functional Forms for Profit and Transformation Functions," *Journal of Economic Theory*, 6, pp. 284-316.

Federal Trade Commission, Bureau of Economics (1982) "Annual Line of Business Report, 1976", May.

Fuss, M. and D. McFadden (1978) *Production Economics: A Dual Approach to Theory and Applications*, North Holland - Elsevier.

Griliches, Z. (1979) "Issues in Assessing the Contributions of Research and Development to Productivity Growth," *Bell Journal of Economics*, v. 10, no. 1.

Hall, B.H., Z. Griliches, and J.A. Hausman (1983) "Patents and R&D: Searching for a Lag Structure," National Bureau of Economic Research, Working Paper No. 1227, Cambridge, MA.

Hausman, J.A. (1978) "Specification Tests in Econometrics," *Econometrica* 46(6), 1251-1271.

Hausman, J.A., B.H. Hall, and Z. Griliches (1984) "Models for Count Data with an Application to the Patents - R&D Relationship," *Econometrica*, 1984.

Hausman, J.A. and W. E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49(6), 1377-1398.

Meta Systems Inc (1982), *Development of a Baseline for Use in Assessing Impacts of the Toxic Substances Control Act (TSCA) on the Chemicals Industry*, prepared for the Economics and Technology Division, Office of Toxic Substances, U.S. Environmental Protection Agency, September 1982.

Muth, John (1961) "Rational Expectations and the Theory of Price Movements," *Econometrics*.

Pakes, A. (1983), "Aggregation Effects and Errors in Variables," *Review of Economics and Statistics*, 1983.

References
(continued)

Pakes, A. and Z. Griliches (1980) "Patents and R&D at the Firm Level: A First Report," Economics Letters 5, pp. 377-381.

Pakes, A. and M. Schankerman (1984), "An Exploration into the Determinants of Research Intensity," Z. Grilliches, ed., R&D, Patents and Productivity, University of Chicago Press, 1984.

Pakes, A. and M. Schankerman (1984), "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in Z. Griliches ed., R&D, Patents and Productivity, University of Chicago Press, 1984.

Schankerman, M. (1981), "The Effects of Double-Counting and Expensing on the Measured Return to R&D," Review of Economics and Statistics, 1981.

Scherer, F.M., (1984) "Using Linked Patent and R&D Data to Measure Inter-Industry Technology Flows," in Z. Griliches, ed., R&D, Patents and Productivity, University of Chicago Press, 1984.